



THE UNITY OF A PERSON

PHILOSOPHICAL PERSPECTIVES

Edited by
Jörg Noller



The Unity of a Person

What constitutes personhood? How are persons and their bodies related? What is the relationship between personhood and value?

The Unity of a Person: Philosophical Perspectives explores the current debates surrounding the philosophy of personal identity and offers a fresh approach to this important topic. It is original in bringing together three approaches to personal identity that are traditionally treated separately: the metaphysical, the phenomenological, and the social. By examining these three areas this volume establishes connections between the underlying metaphysical issues surrounding personal identity and the specific forms of personal existence such as self-consciousness, action, and normativity. Topics discussed include personhood and animalism, process ontology, self-identity over time, sociality and personhood, and the normative status of personhood.

With chapters by an outstanding international roster of contributors, this collection will be of great interest to those studying personal identity and the nature of the self in metaphysics, philosophy of mind, and phenomenology.

Jörg Noller is lecturer in philosophy at the University of Munich, Germany. He wrote his dissertation on the problem of freedom after Kant, and his habilitation on personal life forms. He spent research stays in the USA at the Universities of Notre Dame, Chicago, and Pittsburgh. His main areas of research are personal identity, freedom of the will, philosophy of digitalization, and German idealism.

The Unity of a Person

Philosophical Perspectives

**Edited by
Jörg Noller**

First published 2022
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
605 Third Avenue, New York, NY 10158

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2022 selection and editorial matter Jörg Noller; individual chapters, the contributors

The right of Jörg Noller to be identified as the author of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

Names: Noller, Jörg, 1984– editor.

Title: The unity of a person : philosophical perspectives /
edited by Jörg Noller.

Description: Abingdon, Oxon ; New York, NY : Routledge,
2022. | Includes bibliographical references and index.

Subjects: LCSH: Self (Philosophy) | Identity (Philosophical
concept) | Ontology.

Classification: LCC BD438.5 .U55 2022 (print) |

LCC BD438.5 (ebook) | DDC 126—dc23

LC record available at <https://lcn.loc.gov/2021014649>

LC ebook record available at <https://lcn.loc.gov/2021014650>

ISBN: 978-0-367-72283-8 (hbk)

ISBN: 978-0-367-72286-9 (pbk)

ISBN: 978-1-003-15419-8 (ebk)

DOI: 10.4324/9781003154198

Typeset in Bembo
by codeMantra

Contents

<i>Notes on contributors</i>	vii
Introduction JÖRG NOLLER	1
PART 1	
Ontology	3
1 Personhood and the relation between metaphysics and value ERIC T. OLSON	5
2 The Hybrid Account of personal persistence BENJAMIN CURTIS AND HAROLD NOONAN	23
3 Persons, animals, and persistence PAUL SNOWDON	40
4 Personal identity – a process account GODEHARD BRÜNTRUP	52
5 What are persons?: Attempt at a normatively neutral ontological account THOMAS BUCHHEIM	67

PART 2

Subjectivity 79

- 6 Personal identity and the transfer of commitment** 81
ERASMUS MAYR

- 7 Biological subjectivity: Processual animalism as a unified account of personal identity** 100
ANNE SOPHIE MEINCKE

- 8 Coherence in the self-pattern** 127
SHAUN GALLAGHER

PART 3

Intersubjectivity 147

- 9 Animals in person space** 149
MARYA SCHECHTMAN

- 10 Personal identity and responsibility** 165
KATHLEEN WALLACE

- 11 Collective intentionality, we-identity, and the role of narratives in the constitution of friendship** 185
FELIPE LEÓN

- 12 A person's life: Uniting metaphysics and value** 207
JÖRG NOLLER

- Index* 227

Notes on contributors

Godehard Brüntrup SJ is a Professor of Philosophy at the Munich School of Philosophy. His research interests include metaphysics, philosophy of mind, philosophy of language, analytic philosophy of religion, philosophy of psychology, and the history of analytic philosophy.

Thomas Buchheim is a Professor of Philosophy at the University of Munich. Systematically, his research interests include metaphysics and ontology, especially the problem of freedom and the concept of life. Historically, he focuses on Aristotle's and Schelling's philosophy.

Benjamin Curtis main research is in metaphysics and bioethics, and the intersection between the two. After obtaining his PhD Dr Curtis taught for seven years in the philosophy department at the University of Nottingham before joining Nottingham Trent University.

Shaun Gallagher is the Lillian and Morrie Moss Chair of Excellence in Philosophy at the University of Memphis. His research interests include phenomenology and the philosophy of mind, philosophical psychology, embodiment, intersubjectivity, hermeneutics, and the philosophy of time.

Felipe León is an Assistant Professor at the Center for Subjectivity Research at the University of Copenhagen. He holds MA degrees in Philosophy from the National University of Colombia and the University of Copenhagen, and a PhD degree in Philosophy from the University of Copenhagen. León's primary research areas are classical phenomenology, social cognition, and collective intentionality.

Erasmus Mayr is a Professor for Practical Philosophy at the Friedrich-Alexander-University (FAU) Erlangen-Nürnberg. He studied philosophy, logic and philosophy of science, and law, in Munich, where he also completed his PhD and his practical law course and worked at the Philosophy Institute as a 'Wissenschaftlicher Mitarbeiter'. He was a Junior Research Fellow in Philosophy at The Queen's College Oxford, and, after some other positions, he joined the FAU in 2015. His main interests lie in philosophy of action, ethics, and metaethics. His publications include *Understanding Human Agency* (OUP 2011, paperback 2018).

Anne Sophie Meincke holds a PhD from LMU Munich (2012) and was a Research Fellow at the Universities of Innsbruck (2011–2014), Exeter (2014–2018), and Southampton (2018–2019) before moving to the University of Vienna with an ‘Elise Richter’ research grant from the Austrian Science Fund (FWF) (2019–). Meincke’s main areas of research are metaphysics, the philosophy of mind and action, and the philosophy of biology. Her recent publications include the two edited volumes – *Dispositionalism: Perspectives from Metaphysics and the Philosophy of Science* (Springer: Synthese Library, 2020) and *Biological Identity: Perspectives from Metaphysics and the Philosophy of Biology* (with J. Dupré; Routledge 2020). Meincke is an elected member of the Young Academy of the Austrian Academy of Sciences (2020–).

Jörg Noller is a Lecturer in Philosophy at the University of Munich, Germany. He wrote his dissertation on the problem of freedom after Kant, and his habilitation on personal life forms. He spent research stays at the Universities of Notre Dame, Chicago, and Pittsburgh. His main areas of research are personal identity, freedom of the will, philosophy of digitalization, and German idealism.

Harold Noonan was educated at the University of Cambridge, and taught at the University of Birmingham between 1979 and 2004, before which he was a Research Fellow at Trinity Hall College, Cambridge. His principal teaching and research interests include the philosophy of mind and the philosophy of language (especially reference); identity and personal identity; philosophical logic; and the philosophies of Frege, Russell, and Hume.

Eric T. Olson is a Professor of Philosophy at a university he has asked not to be named. His research is primarily in metaphysics and related areas. Topics of particular interest include personal identity, material objects, time, and death. He is perhaps best known for his book *The Human Animal*, which argues for the unorthodox view that we are biological organisms (‘animalism’).

Marya Schechtman is a Professor of Philosophy at the University of Illinois at Chicago, a member of the Laboratory of Integrative Neuroscience, and an Associate Dean in the College of Liberal Arts and Sciences. She received her PhD in philosophy from Harvard in 1988. Her main areas of interest are personal identity, practical reasoning, and bioethics.

Paul Snowdon was a Fellow of Exeter College, Oxford, between 1971 and 2001, and then Grote Professor of Mind and Logic at UCL until retirement in 2014. In a series of articles and a book *Persons, Animals, Ourselves* (OUP 2014) he defended animalism. He also co-edited with Stephan Blatti *Animalism* (OUP 2016) which contains essays discussing that approach. He has published articles on the philosophy of mind and the history of recent

philosophy, and will shortly be publishing *Essays on Perceptual Experience* with OUP, a collection of his papers on perception.

Kathleen Wallace is a Professor of Philosophy at Hofstra University, New York. Her areas of specialization are metaphysics, ethics, personal identity, American philosophy, feminism. She is interested in the metaphysics of personal identity and responsibility.

Introduction

Jörg Noller

The current debate on personal identity is characterized by three different approaches:

- 1 Ontological approaches that deal with the complex substantial constitution of persons and their spatio-temporal structure.
- 2 Theories of consciousness that focus on the mental structure of the subject and the continuity of its memory.
- 3 Agential approaches that focus particularly on the practical identity of the person through moral action, narrative, and biographical forms of life.

While ontological approaches often ignore questions about the practical life form of persons and their normative status, the exact opposite holds for agent-oriented approaches. Here, the practical form of personal existence is analyzed in detail, but its ontological status is often left undiscussed.

In the current debate, Marya Schechtman has criticized this separation between ontological and practical approaches to the identity of persons, and has argued for a “dependence model”, according to which “practical and metaphysical considerations must be woven together in a single account” (Schechtman 2014, 56). Likewise, Eric T. Olson has argued that

no account of our identity has yet been proposed that guarantees [...] the coincidence of what is important in our identity with the actual conditions of our identity. [...] [A]n account of our identity must be ontologically coherent as well as ethically plausible.

(Olson 1999, 165)

The present volume aims to analyze the unity of a person from ontological, subjective, and intersubjective perspectives. In doing so, the relation between the underlying ontology and the specific forms of personal existence such as self-consciousness, action, and normativity shall be further explained.

The contributions to the volume address the following questions:

- What kind of ontological basis (substances, lives, processes) does most justice to the normative dimension of a person?
- How are persons and their living bodies related?
- What are we essentially – human animals or persons?
- How are ontological, subjective, and intersubjective dimensions interwoven in the concept of a person?
- How is personhood related to subjectivity?
- How can we understand the intersubjective and normative reality of persons (the “person space”)?

The aim of this volume is to bring together different recent approaches to personal identity, and to attempt to bridge the gaps between them. Therefore, the three sections of this volume – ontology, subjectivity, and intersubjectivity – are not to be understood as strict distinctions but rather as perspectives on different aspects of the complex phenomenon of personal identity. Most of the papers stem from talks given at the Munich conference on “The Unity of a Person” in April of 2019. This conference took place as part of the research network “Ontologies of Personal Identity”, funded by the German Research Foundation (DFG) (NO 1240/3-1).

Bibliography

- Olson, Eric T. (1999): “Reply to Lynne Rudder Baker”, in: *Philosophy and Phenomenological Research*, 59(1), 161–166.
- Schechtman, Marya (2014): *Staying Alive. Personal Identity, Practical Concerns, and the Unity of a Life*, Oxford: Oxford University Press.

Part 1

Ontology

1 Personhood and the relation between metaphysics and value

Eric T. Olson

1 Personhood and animality

You and I are people (or as lawyers and philosophers say, ‘persons’). I don’t think anyone disputes this. A few metaphysicians deny that we *exist*, but even they agree that if we *did* exist, we’d be people. No one thinks we’re *non*people.

But we’re not only people. We’re also human beings, friends, parents, or sports fans. And we at least appear to be material things like stones, not immaterial things like angels. We seem to be animals: biological organisms of the animal kingdom. You see an animal in the mirror, and that animal seems to be you: you don’t seem to be some other thing.

These further claims are disputed, especially the last: many philosophers think we’re material things but not animals (e.g. Baker 2000: 96–98, 2001: 159, and the references in note 7). But even if we people really are animals, the *property* of being a person is clearly different from that of being an animal, or even being an animal of a certain biological kind. Some say that being an animal is necessary for having mental properties (Geach 1969: 40, Searle 1992: 90), which would presumably make it necessary for being a person. But no one ever thought it was *sufficient* for being a person.

A few say that being a *human* animal is sufficient. To be a person, they propose, is to be an animal of a kind – *Homo sapiens*, for example – whose typical members have certain special mental powers, such as intelligence and self-consciousness, when mature (Wiggins 1980: 171; see also Snowdon 1996). We are people because we’re animals of that kind. Human embryos and human beings in an irreversible vegetative state also belong to the kind and count as people, despite having no mental powers at all. But a god or an angel, not being an animal, could never be a person no matter what mental powers. Theism – the view that there is a personal and immaterial god – would be self-contradictory. Yet even this unusual view would not make being a human animal *necessary* for being a person: members of other intelligent species – elves, say – could also be people.

By far the most common account of personhood omits the zoological requirement: to be a person is simply to have such mental powers as intelligence

and self-consciousness. (Or more precisely, to be a person at a time is to have those mental powers at that time.) This is expressed in Locke's famous definition of 'person' as 'a thinking intelligent being, that has reason and reflection, and can consider itself as itself' (1975: 335). There could be Lockean people who are not organisms – gods or angels, say – if they could be intelligent and self-conscious. And many human animals – embryos and those suffering from severe senile dementia, for instance – are not Lockean people. Some never are. So being a person in this sense does not entail being an animal, and being an animal, even a human animal, does not entail being a person.

I have no view about what it is to be a person, as opposed to a nonperson. I don't find it an interesting question, and see no point in arguing over it. I *am* interested in the metaphysical and moral status of people, or more specifically of ourselves. But this doesn't turn on the definition of personhood.

Still, there is an important question that can be put in terms of Lockean personhood ('personhood' for short). We have seen that personhood neither entails nor is entailed by being a human animal. But are the two properties compatible? Could something be both a person *and* a human animal? In other words, could a human animal be intelligent and self-conscious?

Well, we appear to *be* animals, in which case the two properties are not only compatible but actually coinstantiated. And even if we ignore this appearance and consider the two properties in the abstract, they *seem* compatible. Nothing about what it is to be a person in the Lockean sense conflicts in any obvious way with what it is to be an animal. Being a person and being an animal seem like being round and being red. They don't seem like being round and being square.

If this is so, it's a pretty safe bet that we actually are animals. If it's *possible* for a human animal to be intelligent and self-conscious, surely many of them actually are. A normal, adult human animal has just the right brain, the right surroundings and history, and the right behaviour in both actual and counterfactual situations to be intelligent and self-conscious if any such animal could ever be. If human animals never have these mental properties, the reason can only be that no such animal, or indeed any other biological organism, could possibly have them.

Being a human animal, then, looks compatible with personhood. And in that case it follows almost inevitably that most human animals actually are people, and thus that we're animals.

2 The metaphysical insignificance of personhood

I'll devote the rest of this chapter to personhood.

Personhood appears to have moral significance: to confer a moral status that nonpeople lack. There is a certain moral status that attaches necessarily to all and only people, by virtue of their being people. Although personhood is a purely psychological property, it is essentially correlated with value-theoretic properties.

This claim comes in different versions. One is that (necessarily) something has this moral status at a time just if it's a person at that time. Another is that something has this moral status at a time just if it's a person at *some* time. (This would mean that human embryos, and human beings who have permanently lost their mental powers by lapsing into a persistent vegetative state, have the same moral status as the rest of us.) A third is that something has this moral status at a time just if it could become, or could have been, a person (if it had a normal, mature nervous system, perhaps), even if in fact it never is one. This would include human embryos that die before acquiring any mental properties, and anencephalic babies born without a cerebrum.¹ The difference between these claims is unimportant for present purposes. I'll focus on the first.

Nonpeople may have a moral status too, but not the one that people have. This thought is reflected in the view that, although a dog can be treated badly, a person can be treated worse. The worst thing you can do to a dog is less bad than the worst thing you can do to a person. This is presumably because people, just by being people, can be harmed in ways that nonpeople cannot be. (Or perhaps because beings that are at some time people, or *could* be people, can be harmed in ways that other beings cannot be.)

But whatever moral significance personhood may have (Lockean personhood, that is), it has no *metaphysical* significance. Or at least this is so if being a person is compatible with being an animal. In that case there are no metaphysical principles that necessarily apply to all and only people. There are no metaphysical laws about people as such, just as there are no physical or biological laws about sports fans as such. People are metaphysically on a par with certain nonpeople – even beings that could never become people. The property of being a person has the same metaphysical status as the property of being a sports fan. This is not to deny that there are important metaphysical principles about ourselves. They're simply not principles about people as such: they apply to us not because we're people, but for some other reason.

Why do I say this? Well, suppose for the sake of argument that we really are animals, as we appear to be. (It doesn't matter if we're not: the same reasoning would apply to any being that was both a person and an animal, and such beings are possible if the two properties are compatible. But suppose we are.)

In that case we're not people essentially. Each of us could exist a time when we lack the mental powers that make us people: in an irreversible vegetative state, for instance. Lapsing into such a state does not destroy an organism and create a new one in its place. In fact we each *do* exist for a time without being a person. Human animals don't come into the world with mental powers, but develop them only after months of growth and development. When we were embryos we were not yet people (in the Lockean sense that concerns us here). And this would be so for any other animal people.

We could even exist without ever being people. An unfortunate accident during your gestation could easily have prevented you from becoming intelligent and self-conscious. But it would not, if you're an animal, have prevented you from existing: it would be too late for that. Personhood is a temporary

and contingent property of us, even if it's the normal condition for animals of our kind. In this respect it's like the property of being a language-speaker: it's normal for a human organism to speak a language when mature, but it's not essential to any particular human organism that it do so. *Some* beings may be essentially people – gods or angels, say – but they would belong to a different metaphysical kind from us.

Our being animals would also imply that personhood does not determine our persistence conditions – the conditions of our identity over time. (The persistence conditions for Fs are given by completing the formula 'Necessarily, if x is an F at time t and y exists at another time t^* , $x = y$ iff... x ... t ... y ... t^* ...'.) There are no persistence conditions that necessarily apply to all and only people (or even to all and only beings that are at some time or could have been people). What it takes for a person to persist depends on what sort of person she is: an animal, a god, an intelligent computer, or what have you.

If there were persistence conditions that necessarily applied to all people, they would presumably have to do with psychological continuity. (The formula would be completed with something like ' $\dots x$ is psychologically continuous in a certain way, at t , with y as it is at t^* ' – perhaps with the addition of a 'non-branching' requirement to deal with those cases where someone is psychologically continuous, at a time, with two beings as they are at another time.) But if we're animals, our persistence does not consist in any sort of psychological continuity. Psychological continuity is neither necessary nor sufficient for a human organism to persist.

It's not necessary, because, as we've seen, human organisms start out as embryos with no mental powers at all, and can exist without them in a vegetative state.

It's not sufficient, because one human animal can be psychologically continuous with another. If your brain were transplanted into my head, the resulting being with your brain and the rest of me would be psychologically continuous with you and not with me. This continuity would be continuously physically realized, and there would be no branching. If *any* persistence conditions necessarily applied to all people, this person would be you with a new body, not me with a new brain. Every philosopher I know of who believes that our persistence conditions derive from our personhood accepts this, whether or not they accept a psychological-continuity account.² Yet the operation would not move an animal from your head to mine. It would not pare down an organism until it was the size of a brain and then give it a new set of parts to replace those it lost. It would simply move an organ from one animal to another. Your brain would cease to be a part of one animal and then become part of another one. For the animal, a brain transplant is like a liver transplant.

This is because there are just two animals in the transplant story: the donor and the recipient. If an animal were to go with its transplanted brain, there would be *four* (Olson 2015: 102–105). One would go with the brain: it would start out looking like you, then have the size and shape of a detached

brain, then look like me. While it was brain-sized, it would not be an organism at all. A second animal would stay behind with an empty head after your brain is removed. It might even be alive. But it would not be the original animal from which your brain was removed, as it, we're supposing, went with the brain. The empty-headed organism into which your brain is transplanted – the one that resulted from removing my brain to make way for yours – would be a third animal. It could not be the animal from which my brain was removed, because it, like the first, would go with its brain when it was removed. That would be a fourth animal – the one originally associated with me. So animal one goes with your transplanted brain; animal two is left over after your brain is removed; that organ is then implanted into the empty cranium of animal three, which was created when animal four went with its own detached brain.

The supposition that psychological continuity is sufficient for human animals to persist would also imply a bizarre account of animal identity over time. Removing your brain from your head would create an empty-headed animal (animal two) that did not exist before. (Otherwise there would have been two animals sharing the same skin until the operation separated them.) And implanting your brain into my head would destroy an animal (animal three): otherwise there would again be two animals (one and three) in the same place at once. It would be metaphysically impossible for a human animal to become smaller by losing its brain, even if this would not cause death, or for a brainless and living animal to become larger by acquiring one.

All this seems to me perfectly absurd.³ Whatever it takes for a human animal to persist, it's nothing to do with psychological continuity. Yet all those who think that there are persistence conditions necessarily true of all people think that some sort of psychological continuity is at least sufficient for a person to persist. I conclude that there are no such conditions.

If we're animals, what it takes for us to persist is what it takes for human animals to persist, even those that are not people. And we don't necessarily share them with all people. When it comes to identity over time, we belong with the dogs and the human vegetables, not with the gods and the angels. What it takes for a person to persist depends on what metaphysical kind of person she is. There is, we might say, no specifically *personal* identity over time.

So if we're animals, being a person is a property we have only temporarily and accidentally. We have the same persistence conditions as things that are not people – human vegetables and presumably dogs – and we could have different persistence conditions from certain people – gods or angels or maybe intelligent computers. All our properties of metaphysical importance – being material things, having different parts at different times, being alive, and being conscious, for instance – are shared with nonpeople. In this regard, being a person is like being a sports fan. And again, this is so if it's even *possible* for a person to be an animal.

3 Baker's account of the metaphysical significance of personhood

Some find all this deeply repugnant. They think personhood *must* have metaphysical significance. Lynne Rudder Baker recognized that this required us not to be animals, and took that to be a powerful objection to the view that we are, 'animalism'. It led her to deny, in fact, that any animal *could* be a person.⁴

Baker thought personhood in the Lockean sense must be metaphysically significant in at least two ways. First, whatever has it must have it essentially. Nothing could be only temporarily or contingently a person. She gave a Lockean account of personhood, defining it as having a 'first-person perspective'. To be a person, in other words, is to have the power of self-consciousness. And she insisted that 'the appearance of a first-person perspective makes an ontological difference in the universe' – that is, a difference in what there is (Baker 2000: 163; see also 17, 218–220, and 2001: 179). For a thing to acquire this power, she said, is for it to come into being, not for it to gain a property it didn't have before, like becoming a sports fan. Nothing could be first a nonperson and then, by acquiring the power of first-person thought, become a person. Likewise, to lose this power is not merely to lose a property, but to cease to exist: nothing could carry on without it as a former person. Otherwise, she said, 'persons as such [would] have no ontological significance' (Baker 2000: 220).

For Fs as such to have ontological significance in Baker's sense – that is, for the property *being F* to have ontological significance – is for it to be impossible for anything to be F at one time and exist without being F at another time. The properties with ontological significance are those that cannot be had temporarily, as being a sports fan can.⁵ Baker found it obvious that personhood must have ontological significance – and ontological significance is a species of metaphysical significance.

Is it obvious? Well, to say that people as such have no ontological significance doesn't sound very nice. But then Baker thinks that sports fans as such have no ontological significance, and that doesn't sound very nice either. We need to remember that 'ontological significance' is a term of art, and that 'Fs as such have ontological significance' means no more and no less than that nothing could be F temporarily.

The claim that *people as such* have no ontological significance does not imply that *people* have no ontological significance. In fact Baker never defined the phrase 'Fs have ontological significance'. We can't define it in the same way as 'Fs as such have ontological significance', like this:

Fs have ontological significance =_{df} necessarily nothing is an F temporarily.

That would lead to a contradiction. You and I are bound to have some property that nothing could have temporarily, even if it's not personhood: being a material thing, say. And of course we have many properties temporarily,

such as being an adult. The definition would imply that material things have ontological significance and adults don't. In that case we'd have ontological significance (by being material things), and at the same time we'd lack it (by being adults). And there is no obvious way of modifying the definition of 'Fs as such have ontological significance' to yield an acceptable definition of 'Fs have ontological significance'. I can't think of any useful definition of the second phrase – nor, to my knowledge, did Baker ever use it.

I can see nothing obvious about Baker's claim that personhood has ontological significance. It's not even very plausible. It implies that no person could survive without remaining a person in the Lockean sense – without being able to think about herself as one being among others. If you were to develop severe senile dementia, you would not cease to be a person and continue existing in a gravely debilitated state. You would literally cease to exist. You needn't fear becoming an addled nursing-home resident unable to recognize loved ones or remember what happened a minute earlier: it's metaphysically impossible. There is no point in trying to visit a relative in such a condition – not because she can no longer benefit from your visit, but because she no longer exists. There may be a resident who looks like your relative, but she was never related to you.

Likewise, the ontological significance of personhood implies that none of us existed before we were people. You were never a foetus. You were probably never even a newborn infant: such beings lack the power of self-consciousness. It's not just that they're not *yet* people: it's impossible for them to *become* people. No being that lacks the mental powers required for personhood could possibly acquire them – otherwise something could be a person temporarily and personhood would lack ontological significance. You and I did not exist until some time after the event we misleadingly call our birth. You might suppose that a bad accident in your infancy could have prevented you from acquiring the mental powers required for personhood. But according to Baker it could only have prevented you from coming into being.

These consequences are unattractive, and no one should accept them without a good reason.⁶ I'll consider Baker's arguments for the ontological significance of personhood in the next section.

Baker thought personhood must be metaphysically significant in a second way too: it must determine the persistence conditions of whatever has it. There must be persistence conditions that necessarily apply to all and only people in the Lockean sense. She recognized that this ruled out our being animals, or even the compatibility of being an animal with personhood. If something could be both an animal and a person, people would not necessarily all have the same persistence conditions: those for human people would be different from those for gods. And some people would share their persistence conditions with nonpeople: with human embryos and human vegetables, for instance. She dismissed any such view as 'not even a contender as an account of *personal* identity over time' (Baker 2000: 124; see also Baker 2001: 179, Johnston 2010: 256).

Her claim is that an account of personal identity must give persistence conditions necessarily applying to all and only people: for people as such. But if we were animals, or even if a person *could* be an animal, there would be no such conditions. This wouldn't mean that there were no persistence conditions for *us*. It would mean only that we don't have those persistence conditions by virtue of being people.

Baker put the view that personhood must be metaphysically significant in these two ways by saying that it must be a *primary kind*. A primary kind tells us what a thing is 'most fundamentally'. It's a property that can only be had essentially and that determines the persistence conditions of the things that have it and only them (Baker 2000: 39f.).

4 Baker's principle of the metaphysics of value

Baker thought that personhood must have metaphysical significance – that it must be a primary kind – because of the connection between metaphysics and value:

Maybe from reading Plato at an impressionable age, I have retained the idea that reality and value go together: What something most fundamentally is should ground what is most significant about it.

(2008: 10)

If being a person has *moral* significance, it must have metaphysical significance too. For a property to have moral significance is for it to make a difference to the moral status of things having it. More precisely, a property has moral significance just if, necessarily, a thing has a certain moral status if and only if it has the property. That's how it is with personhood: there is a moral status that, necessarily, all and only people have; and they have that moral status because they're people. Personhood confers a moral status on its bearers that nonpeople cannot have. The property of being a sports fan, by contrast, may *entail* a certain moral status – sports fans have to be people, and so will have the moral status of people – but sports fans don't have their moral status *by* being sports fans, but by being people. There is no moral status had necessarily by all *and only* sports fans.

As we saw in Section 2, a property can have moral significance in different ways. There are different connections that a property can have to a given moral status. The simplest is perhaps this:

A property P has moral significance *iff*. there is a moral status S such that necessarily a thing *x* has S at time *t* *iff* *x* has P at *t*.

If personhood has moral significance, this implies that a thing has the moral status characteristic of personhood only at times when it's a person. That's what Baker thinks, though only because she denies that anything could be a

person temporarily (making the temporal qualifications in the formula superfluous). But some say that whatever could *become* a person – a human foetus, say – has the same moral status as something that actually is a person. They would say instead

A property P has moral significance *iff* necessarily x has a certain moral status at t *iff* at t x either has or could acquire P.

Other variants are also possible. I'll discuss the simpler version, but my remarks can be adapted to the others.

Baker's claim, then, is that any property having moral significance must be a primary kind: it must be had essentially and must determine the persistence conditions of whatever has it. Call this *Baker's principle of the metaphysics of value*. Given that personhood has moral significance, it follows that there must be persistence conditions that necessarily apply to all and only people, and that whatever is a person must be one essentially: just the claims I found so repugnant in the previous section.

5 Problems for Baker's principle

I doubt whether there are deep and systematic connections between metaphysics and value. This broad statement needs certain qualifications.

There is obviously a connection between metaphysics and metaethics. Metaethics is partly about whether there are ethical properties and facts, which is of course a metaphysical question. I'm not talking about the metaphysics of ethical properties and facts, but about metaphysical questions outside metaethics: those to do with human people, for instance. The point at issue is whether they have a deep connection with substantive normative questions, such as whether there is a moral status necessarily shared by all and only people in the Lockean sense.

Nor do I deny that metaphysical claims outside metaethics can connect with substantive claims about value. It would obviously be important to ethics if theism were true, or solipsism. Metaphysics and value are not completely independent. But I expect the connections to be few – like those between geology and economics.

More specifically, I see no reason to accept anything like Baker's principle of the metaphysics of value. And there are powerful arguments against it. To start with, it's actually inconsistent with her claim that personhood has moral significance. *Consciousness* certainly has moral significance. (By 'consciousness' I mean the property of being conscious, or more precisely the power of consciousness, which we typically have even when we're asleep or under general anaesthesia or in a coma). Conscious beings can suffer, and that gives them a certain moral status, even if it's inferior to the one that derives from personhood. So Baker's principle implies that there must be persistence conditions that necessarily apply to all and only conscious beings. The trouble is

that some but not all conscious beings are people. So if we have the persistence conditions of conscious beings, then those conditions do not apply only to people, but to dogs as well, contrary to Baker's claim that personhood must be a primary kind.

In other words: Baker's principle of the metaphysics of value says that any property having moral significance must determine the persistence conditions of all and only things having that property. Given that personhood has moral significance, all and only people must have the same persistence conditions. Seeing as we are people, we cannot share our persistence conditions with any nonpeople. Given that consciousness has moral significance, however, all and only conscious beings must have the same persistence conditions. Seeing as we are conscious beings, it follows that we must share our persistence conditions with all conscious beings, including those that aren't people. So we can't share our persistence conditions with *only* people. Baker's principle of the metaphysics of value implies that nothing can have two different properties of moral significance (unless those properties are necessarily coextensive). Yet personhood and consciousness appear to be such properties.

No doubt Baker's principle could be amended so as to avoid this problem. Most obviously, we could drop the claim about persistence conditions and say simply that any property with moral significance must be had essentially. There is no inconsistency in saying that we have both personhood and consciousness essentially, or that we have both the moral status characteristic of personhood and that characteristic of consciousness, whereas dogs have only the second.

It would be rather disappointing if Baker's grandiose claim about the connection between metaphysics and value came to no more than this. But even this weaker principle is troubling. It implies that no biological organism could ever be a person: the property of being an organism and the property of being intelligent and self-conscious are incompatible. More strongly, it entails a form of substance dualism.

How? Well, given that personhood has moral significance, the amended principle implies that whatever is a person is a person essentially. What's essential to us is not just having the capacity to *become* intelligent and self-conscious, or belonging to a kind whose normal condition is to be so, but actually having those mental powers. This means not just that nothing could be a person without having them: that's true by the Lockean definition of 'person' and is not a metaphysical claim at all. It means that whatever is in fact a person could not exist at all without the powers of intelligence and self-consciousness.

But no biological organism has these powers essentially. Each human animal starts out as an embryo with no mental properties, and could persist in a vegetative state with none. And we should expect the same to hold for any other animal that could be a person. An organism that is *essentially* intelligent and self-conscious is no more possible than an organism that is essentially able to speak a language.

So the claim that any property with moral significance must be had essentially entails that personhood must always be had essentially, which in turn implies, given that no organism could have the powers of intelligence and self-consciousness essentially, that no organism could be intelligent and self-conscious at all. And if an organism could never be intelligent and self-conscious, it's hard to see how it could have any other mental property.⁷ Having mental properties and being an organism are metaphysically incompatible.

This means that what appears to be a conscious, living thing – living in the biological sense – is really two things: one conscious but nonliving and one living but nonconscious. It's a form of substance dualism. It's not necessarily the traditional Cartesian dualism of mind and matter – the view that mental properties are incompatible with being a material thing, so that conscious beings must be immaterial. In fact, Baker claims that you and I are material things (2001: 159). It is, rather, a dualism of mind and *life*.

Cartesian dualism raises a hard question: Why should being composed of matter be incompatible with thinking or being conscious? What is it about material things that necessarily prevents them from having mental properties? This question must have an answer, yet Cartesians seldom address it. That makes the view mysterious.

Mind-life dualism is even more mysterious. (Mind-life dualism without Cartesian dualism, that is.) If any material thing could ever have mental properties, we'd expect it to be an organism: a human animal, for instance. Mind-life dualism rejects this: it says that some material things *can* think, but not organisms. There would need to be a good explanation for this astonishing fact, and I have yet to see it.⁸

That's why I think we're animals. It seems possible for an organism to have mental properties (especially if any material thing can). In that case it's possible for the human animal writing these words to have them. And if that animal *could* have mental properties, it has them now. But if it has any mental properties at all, then seeing as it shares my brain, my surroundings, my history, and my behaviour in both actual and counterfactual situations, it has the mental properties that I have. Yet it could hardly be a second philosopher in addition to me. It follows that I'm an animal.

6 What matters in identity

So Baker's principle of metaphysics and value – that all properties with moral significance must have metaphysical significance as well – leads all but inevitably to a dualism of mind and life: to the metaphysical impossibility of any organism's having mental properties. Because I cannot accept this, I reject Baker's principle. But even if the principle is true, it can't provide a satisfying alignment of metaphysics and value.

There will remain a gap between our persistence through time and the moral and practical importance of our persistence. We are not momentary

beings: we remain in existence for around eighty years on average (and longer if certain religious beliefs are true). You are the person who began reading this paragraph a moment ago, and not merely someone very like him or her. Persistence is sometimes called ‘identity over time’, but this phrase often leads to misunderstanding, owing to the many meanings of the word ‘identity’. To say that x persists is to say that x exists at more than one time; and to say that x , existing at time t , is identical to y , and existing at another time t^* is just to say that x and y are one and the same – one being and not two.

Our continued existence seems important to us. I’d rather not die today. I want to finish writing this chapter, to see my children grow up, and to spend my retirement in a nice little house with my partner. What I want, it seems, is not merely that *someone* do these things, but that *I* do. As long as my life is good, I want to continue existing for as long as possible. And if my life becomes intolerable, with no chance of improvement, I’d rather not continue existing. Naturally I’d rather not have *anyone* carry on with a miserable life, simply because I hate to see misery. But I’m not entirely unselfish, and if someone has to suffer tomorrow, I’d rather it not be me. Nor are these mere preferences of mine: I seem to have a *reason* to continue existing if my life is worthwhile. It’s *good* for me.

Our persistence seems important in other ways too. Suppose you knew that someone was going to suffer ten hours of agony tomorrow, and someone else was going to have an ordinary day. And suppose you knew that one of these people was you. Other things equal, you’d probably rather have the other person suffer. And again, you would seem to have a selfish or egoistic reason for this preference: that’s the state of affairs that would be better for you.

Likewise, it seems that of all the people living tomorrow, only one of them will be directly responsible for the things you do today. Only one will be obliged to keep your promises. Only one will be entitled to compensation for wrongs done to you. Only one will hold your academic degrees and be the owner of your property: you.

It’s for these reasons and others like them, it seems, that personal identity over time – that is, which people existing at one time are the same as the beings existing at other times – is important to us. This importance is referred to in the literature as *what matters in identity*, or in survival.

But this link between identity over time and what matters in identity – between metaphysics and value – can be challenged. Imagine once again that your brain is transplanted into my head. Two beings result. One will have your brain and the rest of me. He’ll be psychologically continuous with you: he’ll have your beliefs, preferences, knowledge, and personality, not mine, and will remember only your actions. The other will look like you but have an empty head. He or she may be alive, but will have no mental properties, and will of course not be psychologically continuous with anyone. Call them *Brainy* and *Brainless*.

Nearly everyone agrees that in this case Brainy will have what matters to you in identity and not what matters to me. He’ll be responsible for your

actions and not for mine (they're the ones he'll have on his conscience), and will be obliged to keep only your promises (which only he can remember making). You, and not I, have an egoistic reason to prefer Brainy to be comfortable rather than in pain. Brainy's continued existence satisfies your desire to continue living, not mine. And it seems that Brainy's continued existence, supposing it goes well, is intrinsically good for you, and you alone, giving you a reason to want him to survive and flourish.

It seems that Brainy has what matters in identity to you and not to me. And it may seem that this is because he *is* you. But if you and I are animals, he's not you with a new body, but *me* with a new brain. As we saw earlier, the operation does not move an organism from one head to another, but moves an organ from one organism to another, just as a liver transplant does. So if Brainy has what matters to you identity rather than what matters to me, then what matters can come apart from identity. Someone else could have what matters to you in identity. The practical importance that normally attaches to the numerical identity of a person attaches in this case to something else: presumably some sort of psychological continuity.

So much the worse, you might say, for the claim that we're animals (Baker 2008: 13). Because Brainy would have what matters in identity to you, he must *be* you. You would go with your transplanted brain. As a human animal would not go with its transplanted brain, but would remain behind with an empty head, you cannot be an animal. Even though no one is ever actually going to have a brain transplant, we each have a modal property that no animal has, namely being such that we *should* go with our brains if they were transplanted in a way that secured psychological continuity, and thus what matters in identity, between donor and recipient. And whatever has a property that no animal has—even a modal property—is not an animal. This is perhaps the most common objection to our being animals.

But for well-known reasons, this response cannot secure a necessary correlation between numerical identity and what matters in identity. Imagine that your brain is divided, and each half is transplanted into a different head. As before, each of the two recipients – call them Lefty and Righty – is psychologically continuous with you. And if Brainy has what matters to you in the 'single' transplant case, we should expect both Lefty and Righty to have what matters to you in the 'double' transplant. You will have the same reason to care about their survival and welfare that you normally have to care about your own survival and welfare. They will be responsible for your actions. And so on.

Yet they can't both be you. There are two of them: the result of the operation is two people, not one person with two human parts. And one thing cannot be identical to two things that are distinct from each other. If you and Lefty were one and you and Righty were one, Lefty and Righty could not be two: that's an elementary fact about the concepts *one* and *two*. Yet they *are* two. So even if you would be Brainy in the single transplant, you can't be both Lefty and Righty in the double transplant. Yet both Lefty and Righty

would have what matters to you in identity. It follows that we can have what matters in identity without identity itself (Parfit 1971).

Those who believe that we're composed of temporal parts – 'four-dimensionalists' – can say that you *would* survive the double transplant (Lewis 1976). A temporal part of a thing is an object that is just like that thing while the object exists, only temporally shorter. Socrates' nose is a part of him, but not a temporal part, because it's not just like him while it exists. It's too small, spatially, to be a temporal part of him. His first half, if there is such a thing, is a temporal part of him. It eats and drinks and asks awkward questions. According to four-dimensionalists, there are two people even at the beginning of the double-transplant case. Before the operation, they share their temporal parts and are exactly alike, like two roads that share a section in common but diverge elsewhere. When we say 'you' before the operation, we refer ambiguously to both of them. And because both survive the operation, each gets what he or she wants in wanting to survive.

Four-dimensionalism is a controversial metaphysical thesis. But even if it's true, identity will not always coincide with what matters in identity. It implies that Lefty survives the operation and has what matters, after the operation, to herself as she was before it. But then Lefty also has what matters, after the operation, to Righty as *she* was before it (Parfit 1976). Righty has a selfish reason, before the operation, to wish Lefty a long and happy life afterwards. Lefty is responsible, after the operation, for Righty's actions before it – seeing as they're also Lefty's actions. And so on. Nothing could make it the case that Lefty had, after the operation, what matters to Lefty but not what matters to Righty before the operation, because there is no difference between them before the operation or in the causal and psychological relations they bear then to Lefty as she is afterwards.

In fact four-dimensionalists will find it hard to deny that we have what matters without identity even in ordinary cases (Olson 2010). They take us to be composed of *arbitrary* temporal parts: for every period in your life, whether short or long, continuous or discrete, there is a temporal part of you having the same mental powers that you have during that period and existing only then. So there is a temporal part of you that exists from midnight last night till midnight tonight: your 'today part'. It's exactly like you in every respect apart from its temporal extent (and what follows from this).

This doesn't mean that your today part is a person. Call a momentary (or very brief) temporal part of a person a 'stage'. Four-dimensionalists generally say that a person is a thing composed of person stages, each of which is psychologically continuous in some way with every other (Lewis 1976). Or rather, a *maximal* such thing: one that is not a part of any larger one. So your today part is not a person because, although it's an aggregate of person stages, each of which is psychologically continuous with every other, it's not a maximal such aggregate. It fails to qualify as a person simply because it has the wrong temporal neighbours.

Person or not, though, your today part is intrinsically indistinguishable from a person, and should have the moral status of one. And it's plausible to suppose that its desire to exist tomorrow is satisfied by the fact that tomorrow someone will be psychologically continuous with him or her in the right way. Its desire to exist tomorrow is satisfied by your existing tomorrow. That you will have a worthwhile life for years to come seems good for your today part. It doesn't seem tragic that your today part won't be there for it, in the way that it would be tragic if you were going to be run over by a bus at midnight. Likewise, your today part would seem to have a selfish reason to care about what happens to you tomorrow, even though these things won't happen to it. If there could ever be a case of what matters without identity, this would be one.

I'm not aware of any account of our persistence, or of the persistence of beings with moral status, that rules out the occurrence of what matters in identity without identity itself. It will always be possible for someone else to have what matters to you in identity – for the relations of moral and practical importance that you normally bear only to yourself to hold between you and another being. This is why I once said that 'no account of our identity has yet been proposed that guarantees...the coincidence of what is important in our identity with the actual conditions of our identity' (Olson 1999: 165). The conditions of our persistence cannot account for the moral or practical significance of our persistence. What matters is not the metaphysical facts about numerical identity, but something nonmetaphysical facts about psychological continuity, perhaps.

7 How identity might always coincide with what matters

So I doubt whether what matters in identity necessarily coincides with identity itself. Even if we accept Baker's principle of the metaphysics of value and deny that any person could be an organism, a gap will remain between the metaphysical facts about ourselves and the normative facts about ourselves.

What would it take to bridge the gap? It would have to be impossible for any being other than you existing at another time to have what matters to you now in identity, or for you yourself to exist at another time without then having what matters to you in identity now. How could this be the case? Well, the facts about what matters might determine the facts about identity over time: call this the 'value-first' proposal. Or the facts about identity over time might determine what matters: the 'metaphysics-first' proposal. (Or some third set of facts might determine both. I won't explore this thought.)

The 'value-first' proposal is that a past or future being is you, rather than someone else, precisely because she has what matters to you. You are that future being who then has what matters to you in identity, and that past being to whom you have what mattered in identity then, because the facts about identity over time are grounded in the facts about what matters. The metaphysics follows the values.

This claim has an epistemic counterpart: that we can work out which future being is you by discovering which one has what matters to you now; and we can work out which past being is you by finding out which one it is for whom you now have what matters. This follows from the value-first proposal together with the assumption that we can know who has what matters to whom without already knowing who is who, which anyone making the proposal will assume. The thought is that we do our value theory first, then derive its metaphysical consequences. If those consequences go against what our metaphysical thinking would otherwise suggest, so much the worse for metaphysical thinking. If the person who got your transplanted brain would have what matters to you, then she must be you, even if it rules out our being animals and leads to a dualism of mind and life. However metaphysically troubling that may be, we have to accept it because it falls out of the value theory.

But this seems to imply that you must be both resulting people in the double-transplant story. The epistemic claim says we can work out who has what matters to whom before knowing who is who. And it seems, before we consider who is who, that Lefty and Righty, being equally psychologically continuous with you,⁹ each have what matters to you. Whatever ethical thinking tells us that you now have what matters to yourself as you were yesterday would appear to tell us that both Lefty and Righty have what matters to you just after the transplant. It then follows from the value-first proposal that you are Lefty and you are Righty. But one thing can't be numerically identical to two things.

So the value-first proposal can't be right. What matters in identity cannot determine who is who – not, at least, if we can know what matters in identity just by thinking about value.

What about the metaphysics-first proposal, that the facts about identity determine what matters? This would mean that a future being can have what matters to you in identity only because it *is* you, and you can now have what matters in identity to a past being only because you are that past being. Value is subordinate to metaphysics. Our metaphysical thinking tells us who is who, and we must infer from this who has what matters to whom. If the result goes against what our ethical thinking would otherwise tell us, so much the worse for ethical thinking.

Because it's metaphysically impossible for both Lefty and Righty to be you in the double-transplant case, it would follow that they can't both have what matters to you. Or suppose our best metaphysics tells us that mental properties are compatible with biological life. This will almost certainly lead us to accept (for the reasons given in Section 1) that we're animals, and thus to deny (for the reasons given in Section 2) that you would go with your transplanted brain, even though the recipient would be uniquely psychologically continuous with you. It will follow that, despite appearances, this person would not have what matters to you in identity.

That would guarantee the coincidence of what's important in our identity with the actual conditions of our identity, but it would mean abandoning much of our thinking about what matters. It's certainly not the sort of connection between metaphysics and value that Baker had in mind. I doubt whether any account of our persistence conditions can guarantee the coincidence of anything like what *appears* to matter in identity with the actual conditions of our identity.

I can see no reason to make value theory subordinate to metaphysics or vice versa. I'd rather let the two subjects flourish separately.¹⁰

Notes

- 1 This is not the same as the view that personhood in Wiggins' sense has moral significance. The difference is that this view, but not that one, implies that self-conscious beings not belonging to the appropriate animal kind – gods, say – have a special moral status.
- 2 Baker and Swinburne, for instance, deny that our persistence consists in psychological continuity or in anything else, yet never doubt that you would go with your transplanted brain (Baker 2001: 132–141, Swinburne 1997: 147).
- 3 Madden (2016) makes a valiant attempt to defend it, but even he doesn't suppose that there are persistence conditions applying to all and only people.
- 4 Her view was that a human person is 'constituted by' an animal, but numerically distinct from it. She confused matters by saying that we *are* animals in a derivative sense, and that a biological organism can be a person derivatively (2000: 198f., 2001: 173). But she denied that anything that was an animal strictly speaking – 'nonderivatively', as she put it – could possibly be a person strictly speaking. In my exposition of her view, all predicates are nonderivative.
- 5 Perhaps they must also be properties that cannot be had contingently or accidentally: *being at some time a sports fan* cannot be had temporarily, yet presumably lacks ontological significance. And I suppose they must be properties that can be had at all, to exclude cases like *being a round square*.
- 6 And it's not only Baker who accepts them; for references and further objections, see Olson and Witt (2020).
- 7 This is what many materialists actually say: for example, Johnston (2007: 55), Shoemaker (2008), Lowe (2010). Baker says, unconvincingly, that it's metaphysically possible for an organism to be conscious but not self-conscious (2000: 13–16).
- 8 The only such proposal I know of is Shoemaker's (2008).
- 9 They may of course be psychologically different, owing to differences between your cerebral hemispheres; but this is not essential to the story. And both would be psychologically continuous with you to a high degree.
- 10 I am grateful to Karsten Witt for comments on earlier versions.

References

- Baker, L. R. 2000. *Persons and Bodies*. Cambridge: Cambridge University Press.
- . 2001. "Materialism with a human face". In K. Corcoran, ed., *Soul, Body, and Survival*. Ithaca, NY: Cornell University Press: 159–180.
- . 2008. "Big-tent metaphysics". *Abstracta* Special Issue 1: 8–15. <http://abstracta.oa.hhu.de/index.php/abstracta/article/view/73>.

- Geach, P. 1969. *God and the Soul*. London: Routledge.
- Johnston, M. 2007. "Human beings' revisited: My body is not an animal". In D. Zimmerman, ed., *Oxford Studies in Metaphysics* 3. Oxford: Oxford University Press: 33–74.
- . 2010. *Surviving Death*. Princeton, NJ: Princeton University Press.
- Lewis, D. 1976. "Survival and identity". In A. Rorty, ed., *The Identities of Persons*. Berkeley: University of California Press: 17–40.
- Locke, J. 1975. *An Essay Concerning Human Understanding*. P. Nidditch, ed. Oxford: Clarendon Press.
- Lowe, E. J. 2010. "Substance dualism: A non-Cartesian approach". In R. Koons and G. Bealer, eds., *The Waning of Materialism*. Oxford: Oxford University Press: 439–461.
- Madden, R. 2016. "Human persistence". *Philosophers' Imprint* 16 (17): 1–18.
- Olson, E. 1999. "Reply to Lynne Rudder Baker". *Philosophy and Phenomenological Research* 59: 161–166.
- . 2010. "Ethics and the generous ontology". *Theoretical Medicine and Bioethics* 31: 259–270.
- . 2015. "What does it mean to say that we are animals?" *Journal of Consciousness Studies* 22 (11–12): 84–107.
- Olson, E. and K. Witt. 2020. "Against person essentialism". *Mind* 129: 715–735.
- Parfit, D. 1971. "Personal identity". *Philosophical Review* 80: 3–27.
- . 1976. "Lewis, Perry, and what matters". In A. Rorty, ed., *The Identities of Persons*. Berkeley: University of California Press: 91–96.
- Searle, J. 1992. *The Rediscovery of the Mind*. Cambridge: MIT Press.
- Shoemaker, S. 2008. "Persons, animals, and identity". *Synthese* 163: 313–324.
- Snowdon, P. 1996. "Persons and personal identity". In S. Lovibond, ed., *Essays for David Wiggins*. Oxford: Blackwell: 33–48.
- Swinburne, R. 1997. *The Evolution of the Soul*, rev. ed. Oxford: Oxford University Press.
- Wiggins, D. 1980. *Sameness and Substance*. Oxford: Blackwell.

2 The Hybrid Account of personal persistence

Benjamin Curtis and Harold Noonan

1 Introduction

What kinds of changes can persons survive? In particular:

Q1: Can they survive the destruction of their bodies?

Q2: Can they survive an irreversible loss of consciousness?

The philosophical literature contains excellent reasons for answering each of these questions ‘yes’. The main reason in favour of answering ‘yes’ to Q1 is that we react to thought experiments involving the transfer of a person’s psychology into another body as being cases in which the person goes where their psychology does. Call this reason ‘the transplant intuition’. It has been persuasively developed by, e.g., Shoemaker (1963). And the main reason in favour of answering ‘yes’ to Q2 is that there are compelling reasons to think that persons would survive in a permanent vegetative state, e.g., in a coma in hospital. Call this ‘the coma intuition’. It has been persuasively developed by, e.g., Olson (1997).

So, it would be good if we had a theory that enabled us to answer both of these questions ‘yes’. And yet, in the philosophical literature, such theories are thin on the ground. Instead, the vast majority fall either into the Yes-No camp or into the No-Yes camp.

In the Yes-No camp are, for example, Locke (1975), Shoemaker (1963, 1970, 1997, 2011, 2016), Parfit (1971), Perry (1972), and Noonan (2003). They all say that, in some sense or other, psychological continuity is both necessary and sufficient for the persistence of persons. Thus, a person could survive the destruction of their body so long as their psychology continued elsewhere (e.g., in another body). So, their answer to Q1 is ‘yes’. But a person cannot survive if their psychology comes to an end. So, their answer to Q2 is ‘no’.

In the No-Yes camp are, for example, Williams (1956–1957), Thompson (1997), Olson (1997), and Snowdon (1991, 1995). They all say that, in some sense or other, bodily or biological continuity is both necessary and sufficient for the persistence of persons. Thus, a person could not survive the destruction of their living body. So, their answer to Q1 is ‘no’. But a person could

survive an irreversible loss of consciousness if their living body is preserved (e.g., in an irreversible coma in a hospital). So, their answer to Q2 is ‘yes’.

In the first part of this chapter we reject the theories from both camps. Instead, we defend the theory that psychological continuity and biological continuity are *each* sufficient for the persistence of persons, and that their disjunction is necessary. We argue, that is, for a Yes-Yes theory. Because the ‘Yes-Yes’ theory is not a very memorable name, however, we call it the ‘Hybrid Account’ of personal persistence.

In the second part of this chapter we then discuss a recent paper by Olson and Witt (2020). Discussing the arguments they present there, and showing why they fail, will enable us to clarify how the Hybrid Account relates to accounts from the other two camps, and so help to clarify the Hybrid Account further.

2 The Hybrid Account of personal persistence

As a preliminary, we first make it explicit that we think that how we should answer Q1 and Q2 is largely determined by our understanding and use of the concept of a person. We take this concept to be given by our understanding and use of the first-person reflexive pronoun ‘I’. That is, we take the concept of a person that is relevant to debates about personal persistence to be the following concept: the object of first-personal thought. And we take the content of this concept, and thus its extension, to be determined by its use within our linguistic community. This is why we take the transplant and coma intuition so seriously. Considered from the first-person point of view each of us finds the thought ‘*I* would survive’ to be highly plausible in both transplant and coma cases. And this, we take it, gives us strong evidence about how we apply the concept of a person, and specifically about what persistence conditions we associate with it. And so, we take it that this gives us strong evidence for our Yes-Yes Hybrid Account.

So, on our view persons are objects of first-personal thought. But now, if something is an object of first-personal thought then it is a thing that is capable of thinking about itself. It is not required, however, that it is capable of thinking about itself at *all* times at which it exists. All that is required is that it is capable of thinking about itself at *some* time at which it exists. So the term ‘object of first-personal thought’ can be expressed in an extensionally equivalent way as ‘thing that is at some time capable of thinking about itself in the first-person’ or, more neatly, ‘thing that is sometimes capable of first-personal thought’. That is, we endorse the following equivalence:

- X is a person iff X is an object of first-personal thought iff X is sometimes capable of first-person thoughts.

Now, because we respect the coma intuition, we think there *are* persons that are *not* capable of thinking about themselves at the ends of their existences,

i.e., those who fall into unthinking comatose states. So, on our view, persons are things that can lose the capacity for thought altogether. So, if a human person falls into an irreversible unthinking comatose state, it is no longer capable of thinking about itself in the first-person. But, before it fell into that state its first-personal thoughts had as their object a thing (i.e., itself) that includes those later unthinking stages as parts.¹ Similarly, on our view, persons are things that can start off as unthinking things, and only later gain the capacity for self-conscious thought. This is because as well as respecting the coma intuition, we also respect the foetus intuition, i.e., the fact that each of us thinks that it is overwhelmingly plausible that we were once unthinking foetuses that only later developed consciousness, then self-consciousness. So, at the beginning of its existence a human person cannot think about itself (or indeed, about anything else). But it can do so later when it develops and gains certain high-level psychological properties, and what it later thinks about when entertaining first-person thoughts is a thing (i.e., itself) that includes unthinking foetus stages as parts.

With the above understanding of personhood in place, then, we state the problem of personal persistence as follows:

The Problem of Personal Persistence: What changes can things that are sometimes capable of first-person thought survive, and what changes bring about their destruction? (Equivalently: What changes can objects of first-person thought survive, and what changes bring about their destruction?)

In fact, it seems unlikely to us that there is any univocal answer to this question, for as we said above, it is one about the content of first-personal thoughts themselves and is thus determined, as the contents of all thoughts are, by complicated facts about how their constituent concepts are put to use by a linguistic community. And the point here is that there may well be communities of beings (i.e., on other planets) capable of first-personal thoughts who use first-personal concepts in different ways from how we use them here on Earth, and so who associate different persistence conditions with themselves. And in fact, it is difficult to know what to say about non-human animals like dolphins and great apes who appear to be capable of first-personal thought, but who may not associate any clear persistence conditions with themselves at all. For as we have said, our only insight into which persistence conditions we (i.e., human persons) associate with ourselves is given by our intuitions regarding which changes we can and cannot survive, and it is unclear whether *our* intuitions about what changes dolphins and great apes can survive should be taken to determine what changes *they* can in fact survive. That is to say, *our* concept of what changes *we* can and cannot survive determines *our* persistence conditions, but we should not presume that *our* concept of what changes we can survive determines the persistence conditions of *all* persons.

Be the above as it may, we *do* have intuitions about what changes *we* can and cannot survive, and that is enough to be getting along with.

As already mentioned, our intuitions regarding what changes we can survive include the coma, the foetus, and the transplant intuitions. It is because of the first of these that we hold that persons can survive an irreversible loss of consciousness. In effect, we have the intuition that we can survive whatever changes our living bodies can survive, and this gives us good evidence that we can indeed survive those changes. But we also have the intuition that we can survive our consciousness being transplanted into another body.² In effect, we have the intuition that we can survive the loss of our particular living bodies, so long as our consciousness continues elsewhere, and this gives us good evidence that we can in fact survive such a change.

It might be objected at this point that there are possible cases in which both of the following occur: our living bodies lose consciousness but nonetheless survive, *and* our consciousness continues in some other living body. And in this case, it might be said, we cannot survive twice over, for then each of us would be two things and not one. But the structure of this problem is identical to the familiar problem of branching that everybody faces, no matter which theory of personal persistence they adopt. Those, such as animalists, who think that human persons survive if and only if their living bodies survive face the problem that it is possible for human bodies to split amoeba-like such that two bodily continuers result. And those, such as psychological theorists, who think that persons survive if and only if their consciousness continues elsewhere face the problem that a person's consciousness can continue in two separate bodies. One standard response to this problem, and one that we adopt, is to endorse the multiple occupancy thesis, viz. the thesis that before the branching takes place there are two persons present. (See, e.g., Robinson 1985.)

We have said our use of the first-personal concept 'I' determines the persistence conditions of human persons (i.e., us). This requires that those persistence conditions can be formulated as conceptual truths, i.e., as true *de dicto* necessities. And indeed, they can be so formulated. The problem of personal persistence is often put as if it is a problem about identity, i.e., as follows:

What is the relation R that satisfies the following schema:

Person x at t1 = person y at t2 iff person x at t1 bears R to person y at t2

But this problem is not really one about identity at all, but rather one about what changes persons can and must survive. The schema can be split into its two component necessary (right-to-left) and sufficient (left-to-right) conditions, and the problem restated in two English questions as follows:

- 1 Under what conditions can a person at one time be identical with a person at another?
- 2 Under what conditions must a person at one time be identical with a person at another?

And these questions are equivalent to the following questions, now put in terms of the truth of *de dicto* necessities (remembering that a person is thing that at some time thinks about itself in the first-person):

1★. Which relations R satisfy the following schema: Necessarily, if x is a person, then if x exists at t1 and t2, then $R(x, t1, t2)$?

2★. Which relations R satisfy the following schema: Necessarily, if x is a person, then if x exists at t1, then if some person y exists at t2, and $R(x, t1, y, t2)$, then $x=y$?

Specifying a relation that satisfies 1★ gives us a sentence that expresses a *passing-away* condition for persons, i.e., one that places a constraint on how persons can vary across time. In ordinary English, it tells us that persons cannot survive certain changes. Specifying a relation that satisfies 2★ gives us a sentence that expresses a *preservation* condition for persons, i.e., one that specifies how persons must vary across their temporal extents. In ordinary English, it tells us that persons must survive any changes so long as certain continuities nonetheless obtain. So, if we exhaustively specify every relation R that satisfies 1★ and 2★ we thereby specify precisely how persons cannot and can vary across their temporal extents. In ordinary English, we say precisely which changes persons can and cannot survive. And as all of this is specified in terms of *de dicto* necessities, all of this is laid down as being a matter of conceptual truth. Put in such terms, it is simply part of our concept of a person that persons can and cannot survive such changes.

Now, those who endorse the animalist view and its ilk (i.e., human persons survive if and only if their living bodies do) endorse (very roughly) the following view about how to specify relation R:

- Necessarily, if x is a person, then if x exists at t1 and t2, then x has a body at t1 and a body at t2 that are linked by chains of spatiotemporal/biological continuity and/or connectedness.
- Necessarily, if x is a person, then if x exists at t1, then if some person y exists at t2, and x at t1 has a body that is linked by chains of spatiotemporally/biologically continuity and/or connectedness to the body of y at t2, then $x = y$.³

And those who endorse the psychological account (i.e. that persons survive if and only if their psychology continues somewhere) endorse (very roughly) the following view about how to specify relation R:

- Necessarily, if x is a person, then if x exists at t1 and t2, then x has a psychology at t1 and a psychology at t2 that are linked by chains of psychological continuity and/or connectedness.
- Necessarily, if x is a person, then if x exists at t1, then if some person y exists at t2, and x at t1 has a psychology that is linked by chains of psychological continuity and/or connectedness to the psychology of y at t2, then $x = y$.

Our point is simply that our concept of a person, as illuminated by the transplant, fetus, and coma intuitions, strongly suggests that neither of these accounts is correct, and that instead we ought to adopt the following view (again, very roughly stated):

- Necessarily, if x is a person, then if x exists at t_1 and t_2 , then *either* (i) x has a body at t_1 and a body at t_2 that are linked by chains of spatiotemporal/biological continuity and/or connectedness *or* (ii) x has a psychology at t_1 and a psychology at t_2 that are linked by chains of psychological continuity and/or connectedness.
- Necessarily, if x is a person, then if x exists at t_1 , then if some person y exists at t_2 , and either (i) x at t_1 has a body that is linked by chains of spatiotemporally/biologically continuity and/or connectedness to the body of y at t_2 , *or* (ii) x at t_1 has a psychology that is linked by chains of psychological continuity and/or connectedness to the psychology of y at t_2 , then $x = y$.

This is the Hybrid Account. It is more complicated than animalism and the psychological account, for sure. But it is perfectly coherent. And given the strength of the transplant and coma intuitions, this does appear to correctly capture the persistence conditions associated with our concept of a person, and so give the correct answer to the problem of personal persistence. So, we endorse it.

We finish this section by considering two theses that have recently been discussed by Olson and Witt in their (2020) paper. This will lead into the next section where we discuss the arguments contained within that paper in more detail. The two theses are these:

Weak person essentialism: necessarily, if something is a person at a time, then there is no time at which it exists but is not a person.

Strong person essentialism: necessarily, if something is a person at a time, then necessarily there is no time when it exists but is not a person.

On our understanding of the concept of a person, the weak thesis is a trivial truth. It says that, necessarily, if something is sometimes capable of first-personal thought, there is no time at which it exists but is not sometimes capable of first-personal thought. To see that this is trivial take F to be the property of *being a thing that is sometimes capable of first-personal thought*. Then this thesis has the obviously valid form: if something is sometimes an F , then it is always sometimes an F .

As for the strong thesis, we reject it outright. It entails that nothing that is a person in the actual world could be a non-person in some other world. But note that on our view each of us was once an unthinking fetus. But, as seems clear, each of us could have died in the womb before developing consciousness or self-consciousness. As such, each of us could have died before we developed the capacity for first-personal thought. In such a case we

would never have thought about ourselves in the first-person, and so would have failed to be persons. So, each of us is a thing that is a person, but each of us could have failed to be a person. Thus, strong person essentialism is false.

So, on our view weak person essentialism is trivial, and strong person essentialism false. And this is perfectly consistent with our Hybrid Account of the persistence conditions of persons. But it is important to note that this is not a feature that is unique to our view. In fact, Shoemaker's well-known psychological account of personal persistence also has this feature. He too accepts that weak person essentialism is trivially true and rejects strong person essentialism, and this is consistent with the psychological account of the persistence conditions of persons given above. That we say these things would no doubt come as a surprise to Olson and Witt themselves, because they claim that Shoemaker (amongst other psychological theorists) holds a particular version of *both* weak and strong person essentialism called '*Lockean person essentialism*', and that this is *inconsistent* with the psychological account of personal persistence. However, these claims of Olson and Witt are false. It will prove instructive to see why, as it will enable us to spell out more precisely how our Hybrid Account compares with both animalism and the psychological account.

3 Olson and Witt's arguments

In '*Against person essentialism*' (2020) Olson and Witt identify a view they call '*Lockean person essentialism*'. They make three claims about it: first, it is a false unquestioned dogma. Second, it is inconsistent with the standard psychological account of personal persistence, even though the two are almost always held together. And third, there is no easy '*intermediate*' position between the inconsistent combination of Lockean person essentialism with the psychological account on the one hand, and animalism on the other (the two '*extremes*' (Olson and Witt 2020: 17)).

In this section we argue that:

- 1 Lockean person essentialism *is* false. But it is not an unquestioned dogma. Most contributors to the debate about personal identity over time, self-described psychological theorists included, need not be understood as accepting it. The evidence for its general acceptance is rather explained by the fact that psychological theorists accept the definition of a person we have given above.
- 2 Lockean person essentialism *is* inconsistent with the psychological account, or, as Olson and Witt cautiously say, almost every possible version of that view, and every version that has ever actually been held by anyone. But:
- 3 An intermediate view between the two '*extremes*' which denies Lockean person essentialism is not difficult to come by. The standard psychological account as espoused by Shoemaker is exactly such a view.

As should be clear from Section 2, however, we are not defending the psychological account; we think it false for the reasons already given. But it is useful to work through Olson and Witt's arguments in order to see why the Hybrid Account we offer is the *real* intermediate view, rather than the view Olson and Witt identify as such (which is really just the standard psychological account).

Now, though we think that what Olson and Witt call 'Lockean person essentialism' is false, as already indicated in Section 2, we think that weak person essentialism is true (in fact, trivially true) and consistent with the standard psychological account of personal persistence. And it is, we think, merely weak person essentialism that the typical psychological theorist (including Shoemaker) adheres to. Why is this important? First, it means that there is not, as Olson and Witt claim, a fundamental and long-unnoticed incoherence in the standard psychological account. But second, and more importantly, as we will explain in more detail later, getting clear about this enables us to see that Olson and Witt set up the debate between the psychological theorists and animalists in a way that excludes what may be the most promising option, viz. our very own Hybrid view. Olson and Witt think the only options are: (i) the inconsistent conjunction of Lockean person essentialism with the psychological account, (ii) animalism, and (iii) an 'intermediate view' that they think is unsatisfactory. But, as we have said, we will argue that the 'intermediate view' they describe is just the standard, consistent, psychological account itself. Because of the way they set up the debate they miss the possibility of a position which rejects both the standard psychological account and animalism. This middle way, the Hybrid view that we have already outlined, stands between *their* 'intermediate view' and animalism, and is the *real* intermediate view. It retains the psychological theorists' thought that psychological continuity suffices for personal identity but also the animalist view that it is not necessary.

So, now we turn to what Olson and Witt say about all of this. Olson and Witt derive Lockean person essentialism from person essentialism, the definitions of which we reproduce for the sake of clarity:

Weak person essentialism: necessarily, if something is a person at a time, then there is no time at which it exists but is not a person.

Strong person essentialism: necessarily, if something is a person at a time, then necessarily there is no time when it exists but is not a person.

What these theses amount to depends upon what is meant by 'person'. Olson and Witt suppose that psychological theorists typically define 'person' much as Locke does, i.e., as 'a thinking, intelligent being with reason and reflection that can consider itself as itself, the same thinking thing, in different times and places'. Indeed, we do too, but we think that this is best captured by the notion of a thing that is *sometimes* capable of first-personal thought. However, Olson and Witt think that it is best captured without the temporal

qualification, i.e., simply by the notion of a thing that is capable of first-personal thought. So, by plugging this definition of a person into the two person essentialism theses they define a weak and strong version of Lockean person essentialism thus:

Weak Lockean person essentialism: necessarily, if something is capable of first-personal thought at some time, then there is no time at which it exists but is not capable of first-personal thought.

Strong Lockean person essentialism: necessarily, if something is capable of first-personal thought at some time, then necessarily there is no time when it exists but is not capable of first-personal thought.

It is these two theses that Olson and Witt claim are standardly held by psychological theorists, despite the fact they are inconsistent with the standard psychological persistence conditions for persons.

Now, with regard to strong Lockean person essentialism, we have little to say. We do not know why Olson and Witt think that this thesis is standardly held by psychological theorists. Maybe there are some psychological theorists who hold it to be true, but we know of no evidence to suggest that, e.g., Shoemaker holds it to be true, and his account is a paradigm psychological account if any is. Certainly, Olson and Witt provide no evidence for the claim. But, at any rate, we need not quibble about that, because it is really weak Lockean person essentialism that does the work for Olson and Witt, for most of their arguments are aimed at showing that even the weak version is false and inconsistent with the psychological account. And, in addition, they do provide evidence that the weak version is held by paradigm psychological theorists like Shoemaker. So, in what follows, we focus, as Olson and Witt do, on the weak thesis.

Psychological theorists are, of course, committed to saying that you (if you are a person) were never an early stage (unthinking) foetus and will never fall into a permanent unthinking comatose state. But Lockean person essentialism implies that you were never a late stage conscious foetus either, and will never suffer from severe (stage 3) dementia where you retain consciousness but lose the capacity for first-personal thought. And, Olson and Witt argue, this is in fact inconsistent with the psychological account (or almost every possible version of that view and every version anyone has actually held). For you, as you are now, fully self-conscious and capable of first-personal thought are psychologically continuous with the late stage foetus which, though conscious, did not possess the capacity to think about itself in the first-person (any more than a dog does). And if severe dementia lies in what anyone would call *your* future, you are now psychologically continuous with a human being which will be, though conscious, no more capable of first-personal thought than a dog.

In more detail, Olson and Witt argue for this as follows. Psychological continuity is defined via psychological connectedness. Psychological

connectedness is explained as causal dependence between later and earlier psychological states. But when you *first* became self-conscious you were psychologically connected in this sense with the foetus (or infant) as it was just previously. You inherited memories, preferences, and other mental states. Similarly, the devastation of severe dementia is not complete. The resultant being will be in mental states causally dependent on those you were in before your loss of higher consciousness. There will still be connections and hence continuity.

Of course, Olson and Witt point out, we could define a notion of connectedness that avoided this result. We might define connectedness, for example, only in terms of causal dependence between mental states that require self-consciousness and hence ensure that no self-conscious being is psychologically continuous with a late stage conscious, but not self-conscious, foetus, or with someone suffering from severe dementia, simply by ignoring any causal dependencies between any non-self-conscious mental states. But no one has done this, and Olson and Witt explain, doing so would have a consequence that ‘only a tough-minded metaphysician could seriously believe’ (2020: 4): when your friend Sally is affected by severe dementia, though still having many psychological states causally dependent in the right sort of way on psychological states she had before, when she was self-conscious, in fact, on this story, she is dead. She has the same status she would have had if she had died and been cremated – the nursing home resident is a being you have never seen before.

This argument is entirely convincing. Weak Lockean person essentialism is inconsistent with the psychological account. But all that this shows is that psychological theorists, to be consistent, should reject it. And we think it is clear how they should do so. They should simply reject the definition of a person that Olson and Witt attempt to foist upon them and instead accept the definition of a person we have given, and thus accept weak person essentialism in the form that we accept it, i.e., as a trivial truth.

In fact, we think that this is precisely Shoemaker’s position. To see this, consider that Shoemaker notes that the transplant intuition is also plausible if we suppose it is dogs that have their psychologies transferred into new bodies rather humans. As Shoemaker says (2011: 370):

There is no word for dogs that has a definition similar to Locke’s definition of person, one that would lead one to expect a psychological account of their persistence conditions. Still brain transplants involving such creatures are imaginable and the conditional proposition that [psychological theorists] are committed to in the case of persons, namely that if such a transplant resulted in full psychological continuity between donor and recipient the transfer would be mental-subject-preserving, seems plausible here as well ... if the recipient ... recognizes the owner of the original dog ... fawns on them ... knows its way around their house ... digs for bones where the donor buried them ... it would be hard to deny that it is the old dog in a new body.

But how can we say this whilst acknowledging that dogs are animals, as we must? Shoemaker explains:

... we can say that dogs are animals in the same sense as persons are but can deny that their persistence conditions are biological rather than psychological, and so can deny that they are biological animals. Dogs, like persons, will coincide with biological animals, but will not be identical with them. The same will be true of chimpanzees ...

What this overwhelmingly suggests is that Shoemaker thinks of psychological continuers as a general sort of thing under which persons (e.g., the minded beings embodied in human bodies) fall, but under which also fall other minded beings that stand to their bodies in the same way that persons stand to their bodies, but for which we have no specific term (the beings embodied in dog bodies, chimpanzee bodies, etc.). For all of which, he strongly suggests, psychological continuity suffices for survival. And so, on this suggestion, we should think of our lives as possibly extending backwards and forwards in time to periods when we lack the capacity for first-personal thought, which other individuals of the sort *never* possess. And thus, it seems, he thinks of persons in a very similar way to the way we define them, viz. as being things that *sometimes* possess the capacity for first-personal thought.

Nevertheless, Olson and Witt argue there is evidence in the writings of Shoemaker, and indeed in basic textbook level presentations of the problem of personal identity, that Lockean person essentialism is presupposed, though inconsistent, with the psychological account. But this is not so; the evidence they point to is consistent with the assumption that what the writers in these debates are concerned with is only the persistence conditions of persons as *we* have defined them, and that the only essentialist thesis being presupposed, if any is, is the trivial version of weak person essentialism that we have outlined and that we ourselves are committed to.

One passage from Shoemaker (2011: 360) that they point to, and in which they say he explicitly endorses Lockean person essentialism, reads as follows. After giving Locke's definition of a person, Shoemaker writes:

Certainly, we conceive of persons as creatures that have, and in some sense necessarily have, mental or psychological properties. Of course, an animalist can hold a version of this view; when the person becomes a human vegetable, entirely devoid of mentality, the animalist can say that it ceases to be a person but does not cease to exist. On such a view Locke's definition only gives us the nominal essence of persons. One might think that the nominal essence reading is recommended by Locke's claims about the unknowability of real essences. But the rest of what Locke says in this chapter [Chapter xxvii of Book II of the *Essay*] goes better with the view that it is necessary *de re* of persons that they are beings that satisfy his definition And that, suitably qualified, seems to me an intuitively plausible view.

But does Shoemaker endorse Lockean person essentialism here? Hardly. His comments are intended simply to signal that, for the psychological theorist, the *thing* that is a person ceases to exist when it loses consciousness, whilst for the animalist the *thing* that is a person does not cease to exist. This is the difference Shoemaker wants to stress by contrasting real with nominal definitions, and the *de re* with the *de dicto*. But what he says here is perfectly consistent with his holding the view that persons are things that are sometimes capable of first-person thought. To say that his words contain an explicit commitment to Lockean person essentialism, as Olson and Witt define it, is to push too hard.

Olson and Witt also cite a standard textbook statement of the problem of person identity as evidence that Lockean person essentialism (they actually just write ‘person essentialism’ but they cannot mean our trivial weak person essentialism) is commonly presupposed:

The problem of personal identity over time is the problem of giving an account of the logically necessary and sufficient conditions for a person identified at one time being the same person as a person identified at another.

(Noonan 2003: 2)

This, of course, presupposes that it makes no sense to talk of a thing’s being a person at one time and not at another; it presupposes that personhood need not be temporally qualified (Olson and Witt 2020: 2). But it does not presuppose Lockean person essentialism. Again, it is consistent with ‘person’ meaning ‘thing that is sometimes capable of first-personal thought’.

Olson and Witt quote another passage from Noonan (2003) as evidence of Lockean person essentialism being presupposed:

[W]hen philosophers speak of the problem of personal identity they do not use “person” in the sense of “human being”. Rather they use it in the sense introduced by Locke.

(Noonan 2003: 8)

But once more here, the wording, though vague, does not necessarily presuppose what Olson and Witt claim. The thought expressed here is this: when philosophers use the word ‘person’ they do not use it as a mere synonym of ‘human being’, for they allow the conceivability of non-human persons (e.g., intelligent extra-terrestrials and angels); rather, they follow Locke in thinking that the various capacities he mentions lie at the core of personhood. But this too is compatible with ‘person’ meaning ‘thing that is sometimes capable of first-personal thoughts’. For, of course this definition does contain essential reference to the Lockean capacities.

In relation to the above, Olson and Witt also consider how to reformulate the traditional statement of the problem of personal identity *without* the

presupposition of Lockean person essentialism. But, in fact, no reformulation is needed if ‘person’ is read as we recommend. Similarly, if we read in this way the typical formulation of the psychological-continuity account they give, i.e.,

Necessarily, if a person x exists at one time and a person y exists at another, x is y iff x is in some way psychologically continuous, at the first time, with y as it is at the other time.

(Olson and Witt 2020: 4)

this needs no reformulation either, as it gives us exactly what we want from an account of personal identity over time.⁴ It tells us what changes a thing that is sometimes capable of first-personal thought can survive and what changes it cannot survive. It tells us, in fact, two things – precisely those two things that we have said psychological theorists maintain, i.e., that:

- Necessarily, if x is a person, then if x exists at t_1 and t_2 , then x has a psychology at t_1 and a psychology at t_2 that are linked by chains of psychological continuity and/or connectedness.
- Necessarily, if x is a person, then if x exists at t_1 , then if some person y exists at t_2 , and x at t_1 has a psychology that is linked by chains of psychological continuity and/or connectedness to the psychology of y at t_2 , then $x = y$.

So far, then, we have argued that the standard psychological account, as defended in the literature, is one that does not presuppose Lockean person essentialism and that is, rather, entirely consistent with our definition of a person (and thus the triviality of weak person essentialism). Moreover, we have argued that this is the position Shoemaker, a paradigm psychological theorist, holds. However, Olson and Witt do consider what the psychological account *minus* Lockean person essentialism looks like. It is this view that they call ‘the intermediate view’.

According to the so-called ‘intermediate view’, we are fundamentally psychological beings, but in a weaker sense than Lockean person essentialism implies, i.e., in the sense that even if no particular mental powers such as intelligence or self-consciousness are essential to us, *having some mental power or other* is. This is indeed entailed by the standard psychological account, for on that view psychological continuity is necessary for persistence, and so persons cease to exist when their psychology comes to an end. And so, on this view, we cannot exist without some mental state or other. The so-called ‘intermediate view’ therefore entails a version of weak *psychological* essentialism. This can be spelled out as:

Weak psychological essentialism: necessarily, if something is a person at a time, then there is no time at which it exists but is not conscious.

As we have been at pains to emphasise, *contra* Olson and Witt, a commitment to this principle is not part of a novel ‘intermediate’ view that lies between the two ‘extremes’ of the standard psychological account and animalism. Rather, it just is the standard psychological account to which reflection on transplant cases naturally leads us if it leads us to a psychological continuity account at all. Nonetheless, Olson and Witt do provide reasons to reject it, and so perhaps they have a point to make here after all. Whether they do or not depend on their reasons. So, we now consider them. There are four.

First, they point out that this view implies that no biological organisms have mental powers, not even ones physically indistinguishable from us. This *is* a consequence of the standard view that we persons (i.e., beings sometimes capable of first-person thought) are psychological continuers. The biological organism coincident with me was once a mindless foetus. So, it is not a psychological continuer. So, it is not a *person*. So, it is not even now, when physically indistinguishable from me, a being that is capable of first-personal thoughts. And if it does not now possess that mental power it cannot possess any. But if the complex biological organism coincident with me now lacks mental powers, no biological organism (or any that we know of – Olson and Witt 2020: 17) ever has mental powers. However, this is not a new problem for psychological theorists. And, in fact, this consequence is a *feature* of the standard view, not, its defenders like Shoemaker claim, a flaw. The view just *is* that we and other animals are psychological continuers, but we coincide temporarily (and sometimes permanently) with mindless biological organisms, which we might call, using a term of art, ‘biological animals’ (Shoemaker 2011: 370), just as statues coincide temporarily (and sometimes permanently) with hunks of clay, which though coincident are numerically distinct because differing in some properties.

Second, Olson and Witt point out that according to the so-called ‘intermediate view’, acquiring the higher level powers that makes something a person creates no new entity, and losing them destroys none; people as such have no ‘ontological significance’. Again, however, this is not a problem but a *feature* of the view, not a flaw. Olson and Witt cite Baker (2000) as thinking that it *is* a problem, but this doesn’t make it one. And anyway, Olson and Witt themselves effectively respond to Baker:

“ontological significance” is a technical term. Baker’s statement that Fs as such have no ontological significance in fact means nothing more than that Fs are only contingently F. (For lumberjacks as such to have no ontological significance is for them to be only contingently lumberjacks.) What she presents as an unattractive consequence of denying person essentialism is only a restatement of that denial in dyslogistic terms. Its force is purely rhetorical.

(Olson and Witt 2020: 14)

Third, Olson and Witt point out that the so-called ‘intermediate view’ rules out any account of personal identity as such, i.e., any account that applies

only to persons. According to the view, we are psychological continuers and so share our persistence conditions with things that are never self-conscious beings, like dogs. Again, a feature not a flaw, unsurprising in the light of the transplant intuition, even if some (Baker 2000, Gert 1971, Johnston 2010) say that this makes the view not even a contender for an account of personal identity over time. In other words, these people are simply wrong. From the standard psychological theorists' viewpoint saying this is comparable to insisting that there must be an account of the persistence conditions of *musical geniuses* as such.

Fourth and finally, Olson and Witt point out that the intermediate view is incompatible with Parfit's influential account of personal identity, that a person persists for a day only if she exhibits a certain number of direct psychological connections. They are right, but the particular view of Parfit's they explain is that 'you persist just if you have at least half as many direct psychological connections ... over a day as there actually are during a day in the life of nearly every person' (Parfit 1984: 206). But a psychological continuity theorist should surely not be regarded as committed to this (charmingly precise) condition for persistence.

So, to summarise, we do not think Olson and Witt's 'intermediate view', i.e., the standard psychological account according to which all persons are psychological continuers, has been shown by Olson and Witt in their (2020) paper to have any problems which were not already familiar.

However, as we have said, we do not accept the psychological account, for we respect the coma intuition. Each of us may one day fall into a permanent unthinking state. Psychological theorists may be able to explain senses in which this statement is true, but the difficulty is that they have also to acknowledge senses in which it is false. But there are no such senses. It is just true that each of us may fall into an unconscious unthinking state. Psychological continuity is not necessary for personal identity.

So, should we be animalists? No. The transplant intuition is compelling. Even if psychological continuity is not necessary for persistence, some form of it suffices. We need a middle way.

Participants in the personal identity debate need not assume that there is an account of the persistence conditions of persons as such. But they must assume that there is an account which is true of *all* persons, even if it is true of other things as well. But, when we think about all the kinds of persons which seem conceivable (human, dolphin, extra-terrestrial, robotic, angelic, divine) is this plausible?⁵ Boethius defined a person as an individual substance of rational nature. Perhaps the persistence conditions of a person depend upon the kind of individual substance it is.

The response of the psychological continuity theorist will be that what makes their account compelling is the transplant intuition, and this also shows that there *must* be a common sort to which all persons belong for which persistence conditions can be given. However, what the transplant intuition makes compelling is only that some form of psychological continuity *suffices* for personal persistence. It is consistent with it that the common sort is one for

which it is not necessary as well. We think that is what must be correct, given the coma intuition. Olson and Witt think that the only options are: (i) the inconsistent conjunction of the psychological account with Lockean person essentialism, (ii) animalism, and (iii) their ‘intermediate view’ (which is really just the standard psychological account). Because of the way they have set up the debate however, they have missed another possibility: a position which rejects both the psychological account (because it conflicts with the coma intuition) and the animalist view (because of its conflicts with the transplant intuition). This middle way – the Hybrid view we outlined in Section 2 – is the *real* intermediate view. It is intermediate between two consistent views. It retains the thought of the psychological account that psychological continuity suffices for personal identity, which is all that the transplant intuition supports, but adds that physical continuity also suffices, which is all that the coma intuition supports. It is, we think, the best of both worlds.⁶

Notes

- 1 We do not mean to commit ourselves here to perdurantism by this talk of parts. We wish, rather, to stay neutral between endurantism and perdurantism, and we think that our view can be formulated adequately on either view (given a generous ontology). But we do sometimes help ourselves to the language of perdurantism for ease of expression.
- 2 Or, perhaps, even into a non-living computer system, or some such. But we ignore this here.
- 3 In fact, this is not quite right, as a tweak is needed to allow for branching. But this can be done quite easily, and we ignore this complication here and in what follows. For more detail on this way of spelling out the debate about personal identity see Noonan and Curtis (2018).
- 4 Setting aside, as Olson and Witt note (2020: 4), ‘complications due to branching’.
- 5 See Noonan (1978: 351).
- 6 See also Noonan (2021) where some of the details of the Hybrid Account have also been worked out.

Bibliography

- Baker, Lynne Rudder 2000, *Persons and Bodies* (Cambridge: Cambridge University Press).
- Gert, Bernard 1971, ‘Personal Identity and the Body’, *Dialogue* 10, 458–478.
- Johnston, Mark 2010, *Surviving Death* (Princeton, NJ: Princeton University Press).
- Locke, John 1975, *An Essay Concerning Human Understanding* (Oxford: Oxford University Press) (2nd ed. first published 1694).
- Noonan, Harold 1978, ‘Locke on Personal Identity’, *Philosophy* 53, 343–351.
- Noonan, Harold 2003, *Personal Identity*, 2nd ed. (London: Routledge).
- Noonan, Harold 2021, ‘Personal Identity and the Hybrid View: A Middle Way’, *Metaphysica* 22(2), 1–21.
- Noonan, Harold, and Curtis, Benjamin L. 2018, ‘The Simple and Complex Views of Personal Identity Distinguished’. In Valerio Buonomo (ed.), *The Persistence of Persons Studies in the Metaphysics of Personal Identity Over Time* (Neunkirchen-Seelscheid: Editiones Scholasticae), 21–40.

- Olson, Eric 1997, *The Human Animal* (New York: Oxford University Press).
- Olson, Eric, and Witt, Karsten 2020, 'Against Person Essentialism', *Mind* <https://doi.org/10.1093/mind/fzaa016>.
- Parfit, Derek 1971, 'Personal Identity', *The Philosophical Review* 80(1), 3–27.
- Parfit, Derek 1984, *Reasons and Persons* (Oxford: Oxford University Press).
- Perry, John 1972, 'Can the Self Divide?', *The Journal of Philosophy* 69(16), 463–488.
- Robinson, Denis 1985, 'Can Amoebae Divide without Multiplying', *Australasian Journal of Philosophy* 63, 299–319.
- Shoemaker, Sydney 1963, *Self-Knowledge and Self-Identity* (Ithaca, NY: Cornell University Press).
- Shoemaker, Sydney 1970, 'Persons and Their Past', *American Philosophical Quarterly* 7(4), 269–285.
- Shoemaker, Sydney 1997, 'Parfit on Identity'. In J. Dancy (ed.), *Reading Parfit* (Oxford: Blackwell), 135–148.
- Shoemaker, Sydney 2011, 'On What We Are'. In Shaun Gallagher (ed.), *The Oxford Handbook of the Self* (Oxford: Oxford University Press), 352–371.
- Shoemaker, Sidney 2016, 'Thinking Animals without Animalism'. In S. Blatti and P. Snowdon (eds), *Animalism: New Essays on Persons, Animals, and Identity* (Oxford: Oxford University Press), 128–142.
- Snowdon, Paul F. 1991, 'Personal Identity and Brain Transplants', *Royal Institute of Philosophy Supplement* 29, 109–126.
- Snowdon, Paul F. 1995, 'Persons, Animals, and Bodies'. In J.L. Bermúdez, A. Marcel and N. Eilan (eds.), *The Body and the Self* (Cambridge: MIT Press), 71–86.
- Thompson, Judith J. 1997, 'People and Their Bodies'. In Dancy (ed.), *Reading Parfit* (Oxford: Blackwell), 202–229.
- Williams, Bernard A. O. 1956–1957, 'Personal Identity and Individuation', *Proceedings to the Aristotelian Society* 57, 229–252.

3 Persons, animals, and persistence

Paul Snowdon

The approach that I favour when thinking about the nature of persons, or as I prefer to say, *our* nature, is, of course, animalism, which identifies each of us with the animal where each of us is, and so holds that each of us persists over time so long as that animal persists. It holds too, putting the point in first-person terms, that there is no more to me, and of course, no less, than what there is to this animal here, where I am. Somethings that were problems about persons, then get transformed into problems about human animals. Still, even though mysteries and questions remain it is reasonable to feel that recognising our identity to a sort of animal represents progress in determining our place in the world – if it is true, of course. I approach the issue with the conviction that the animalist claim, although completely neglected during long periods of discussion, is, once put on the table, not one that strikes us as weird, paradoxical, or preposterous. Rather than rushing to judgement against it, we should try our hardest to unearth the real potentialities of the proposal.

One point to acknowledge, I believe, is that when comparing *yourself* with the *animal* where you are the most striking thing is the extent of how much you and that animal are the same. One would say that the histories of the two things are the same, their physical properties are the same, and so are their psychological properties. This overlap should encourage us to look very hard at grounds being offered to deny the identity. There is also something solid and familiar about animals. With them we are in the realm of the known. You can't just say anything about animals as a kind of thing, and so you cannot about us if we are animals. Further, we can draw on, and gain illumination from the sciences studying animals. In making claims about animals we aren't simply relying on armchair intuitions.

1 Persons

Now, one consequence of the proposal is that it dislodges from centrality in our inquiries that term 'person' which so obsesses philosophers and which, it seems to me, deserves to be called 'a delusion ... and a snare'. It does that because there is no need in discussing the animalist proposal to formulate it

in terms of ‘persons’ and so no need to grapple with analysing that notion. It is, indeed, very hard to determine the basic *logic* of the term ‘person’. Do we become persons at some stage in our development, if so when? And can we cease to be persons? If we do become persons and can cease to be persons, then it does not capture or represent what we fundamentally are and so why is it the central notion for us in thinking about ourselves? Further, what sort of semantic conditions is the term ‘person’ linked to? By ‘semantic conditions’ I mean the conditions for something qualifying as a person. Locke famously selects reason and reflection, and something like self-consciousness (though it is pretty obscure what he has in mind) – but why chose those two, or three, conditions?¹ It is quite simply an obscure term. Maybe this obscurity is reflected in a fact that in the title of the great books in our philosophical tradition their subject matter is fixed, not using the term ‘person’ but the term ‘human’. Locke, just mentioned, is investigating Human Understanding, Berkeley is investigating Human Knowledge, and Hume is investigating Human Nature. None of them signal that they are investigating personal understanding, or personal knowledge, or personal nature. Yet, if what we are is fundamentally persons, why not? So by using ‘person’ to fix the subject matter of a whole branch of philosophy we are centring it on a rather spongy term, and in consequence inviting relatively unanchored speculation, an invitation gleefully accepted, of course. This sponginess is not lessened by seeking illumination in talking of the *life of persons*. One can ask at what stage in the evolutionary development leading to the emergence of *homo sapiens* did they begin living personal lives. Until one has a solidly based answer to that question the notion of leading a personal life is as spongy as the notion of a person. Above all, we should not confuse evident features of the lives lived in advanced western democracies with what personal life *must* contain.

2 The animalism debate

However, every view about our nature has its difficulties and animalism is no exception. To my mind the really difficult type of case for animalism are what used to be called ‘brain transplant cases’, and in this stage of the discussion I shall refer to them in that way. The simplest form of the problem is that it seems that in such imagined cases of brain transplants we go with the brain leaving the animals behind, and so we cannot be the animal, since we have separated. As normally developed this objection appeals, as it is said, to our intuitions when asked to adjudicate a verdict on neutrally described transplant cases. The defence of animalism requires some critical engagement with that sort of case. And then the discussion gets more complex since the replies face difficulties which reflection on those answers generates and those difficulties also have to be faced. Feeling that this area is the most difficult for animalism, but still wanting to work hard to support it, and not wanting simply to repeat the moves that occurred to me when writing about this problem in the past, I want to critically inspect a new, or newish, proposal

about this problem suggested by Rory Madden. The version of his view I shall scrutinise is his paper ‘Human Persistence’ in *Philosophers’ Imprint* Vol 16, No 17, (September 2016) pp. 1–18. I should add that in this discussion I shall have to ignore quite a lot of Madden’s argument, including some of its most interesting claims. My focus is really on pages 1–8 only.

Before starting that engagement I want very briefly to fill in the background to the transplant debate as I see it. In the early days of debates about what was called personal identity the two views which were in conflict were a Lockean style psychological continuity view, and what was called ‘the bodily view’, which said that we, or the person, are or is traceable through bodily continuity. Somewhat later, but well before the emergence of animalism brain-based views of our persistence conditions emerged to complicate the discussion. Shoemaker himself brought the brain transplant case forward to disprove the bodily view. Body theorists at that stage, in a famous contribution by Williams, responded by arguing that the person would be lost because that person could not express itself in the new body simply in virtue of donating the brain, but he also raised problems about fission for Lockeanism, and allied to this other body supporters raised worries about the employment of psychological concepts in the Lockean view. These moves elicited replies, and the initial response of Williams that was supposed to undermine the brain transplant intuition did not win many supporters. Looking back, though, one rather odd feature of the discussion was that brain transplant cases were taken to support pure Lockeanism, but since the supposedly successful transplant cases which are imagined are not ones purely of psychological transfers, the brain being an essential part, it is hard to see why the intuitive verdict on such cases gets one to Lockeanism, rather than to a brain view. What also happened slightly later was the emergence into the debate of the notion of an animal. Opponents of the bodily theory naturally took it that their transplant objection also worked against animalism, so anyone attracted to animalism had to come up with a response to it. But a key question now when we think about these cases is what happens to the *animal* in the case of the transplant. It is clear that animals can lose quite a lot of their body and still survive. So the question is – how much? Everything but their brain? Perhaps more, perhaps less?

3 Madden’s proposal

With that question as the background, what is very interesting about Madden’s approach is that he proposes that if animalism is correct then given a proper analysis of what human animal persistence amounts to it is actually implied that we go with the brain since in our case, given the kind of animal that we are, transplanting the brain *is* transplanting the animal. So the fact that, when we reflect on such cases, it seems that we go with the brain is *not* a difficulty for animalism at all, rather, it *fits* what the general animalist thesis says. Therefore, the so-called intuitions should be viewed not as a difficulty,

as disconfirming, but rather as confirmatory of, or at least consistent with, the animalist view.

The question I am trying to make some progress with here is whether this novel view is a reasonable one. Two things need to be said. First, my own thinking about this approach is at a fairly early stage, so my conclusions are to some extent provisional and not perhaps as clear as I would like them to be. Second, Madden's paper has many virtues and is, I think, bold, important, and deserving of serious attention. With a theory like Madden's we can be lead to object in two ways. One way is to directly oppose, for some reason or other, the verdicts about persistence that he supports. But another way is to scrutinise the basic assumptions and the logical derivations to try to determine whether they can be relied on. This discussion is very much along the second line, and I hope something about how to think (or perhaps, how not to think) about persistence can be learnt in this way.

4 Madden's theory

Madden sets out his approach in Sections 2 and 3 of his paper. Section 2 states his very general assumptions about persistence, and Section 3 applies them to our case. The general approach he favours he sees as in some ways a development of the ideas that Wiggins has proposed and refined over the years, a picture that Madden (following Wiggins) calls 'broadly Aristotelian', and which in Wiggins' exposition Leibniz is also credited with a major role.²

The general category of things that Madden is concerned with he calls 'macroscopic continuants' (p. 4), the examples he gives are 'boulders, cats and trees'. In the famous words of the old quiz we are roughly talking about 'animals, vegetables and minerals'. These examples are meant to fix what the category is. I myself don't want to cavil right at the beginning about that category, except to ask whether this focus is guided by the thought that the theory of the persistence of such relatively large-scale objects is supposed to be different from the theory of persistence of what we might call microscopic objects (presumably the, or, perhaps at least, a contrasting case, and gestured towards by the name 'macroscopic'). On the face of it that would be an uninviting thought. But we can assume that there is no reliance on that and it is left open that the approach can be applied to microscopic things. Another contrasting case would surely be what we might call social objects, such as football teams or orchestras, which are clearly different sorts of things, and their persistence needs to be thought about in a different way. In the list that Madden gives artefacts are absent, and it might be wondered whether they are left off intentionally. In fact the very first case that Madden uses to illustrate the general approach are artefacts, namely toasters.³ So the account *is* meant to fit them.

Now, Madden, approaching things in a broadly genetic way, says that we begin by perceiving these macroscopic objects, and then we, ordinary engaged observers, learn that 'the activities of these entities figure in a range

of law-like generalisations which enable us to systematise and explain external phenomena'. He summarises the conception we end up adopting in the words; 'A macroscopic continuant is ... a locus of law-like activity characteristic of its general kind'. Madden doesn't name this characterisation, but I shall call it 'MC', short for Macroscopic Continuants.

From this characterisation of MC Madden derives the following characterisation of the persistence of macroscopic-objects; 'A continuant of fundamental kind K persists if and only if a sufficient number of capacities for K-characteristic activity are continuously preserved (along a dominant path)'. Madden calls this general principle – 'Persistence'.

Now, one element in this proposal is its inclusion of the notion of a dominant path. This notion is needed given the focus of the discussion, which is to generate verdicts in philosophically disputed thought experiments about our persistence conditions, and, of course, the cases where verdicts need to be grounded are ones where there is some sort of bifurcation in the histories. Part of a human animal – roughly its brain – is removed and has its pathway, and the rest of the body has its pathway (or perhaps pathways). Two things might be said here. The first is that the cases for which the dominant path clause is needed most are not ones that occur in the actual world. Brains are not transplanted, nor are heads. Second, when is the dominant pathway condition supposed to operate in our actual thinking? If a hair drops off my head and on its new pathway it drifts away, is my persistence as the thing without that single hair supposed to be grounded in the application of the dominant path condition?

Next, Madden applies this highly general proposal to the human animal case to get a principle which he calls A-Persistence and which says, 'One of us persists if and only if a sufficient number of capacities for human-animal-characteristic activity are continuously preserved (along a dominant path)'. To this Madden adds the claim that the so-called remnant cerebrum case (that is the case where a cerebrum is removed and sustained in its functioning) continuously preserves (along a dominant path) a sufficient number of capacities for human-animal-characteristic activity. This claim he calls Sufficiency. This amounts to him claiming that the human animal goes with the cerebrum, and so the animalist should agree that such transplants take the animal, and so they need not quibble with the intuition that we go with the cerebrum. What he is at pains to stress though is that his theory also implies that if, say, a human animal has its cerebrum incapacitated but remains alive then that is also the survival of an animal. He is claiming that his theory fits the normal animalist idea that mental functioning is not necessary for human animal survival. This is extremely ingenious and, as far as I know, Madden is the first to develop it properly.

5 Some questions

Just how convincing is this complex package? I want to raise a few questions.

(1) I shall call the first set of problems the problems of *activity*. Note that this notion figures both in the characterisation of MC, and then in the characterisation of persistence for such a type of object. So, what aspects of a something are being picked out by talk of its *activities*?

It is clear that we are not picking out solely actions – things that agents do – actions in a heavy sense. This is obvious because lots of macroscopic objects are not agents in that sense. I am happy to say that animals generally are agents, contrary to what many philosophers might say, but clearly rocks and vegetation are not. (Not that I have tried, here, to spell out that sense.) But is also clear from even the human examples. Madden lists digestion as an activity, but digesting is not an action in the strong sense. It is clearly right to call it an activity.

Second, does an activity require the production of an effect? Does the attribution of an activity require that an effect is produced in some way? Well, as far as I can see from the examples that are given this is not necessary for an activity. Madden mentions a boulder rolling down a slope. It is natural to call that an activity, the activity of changing location in a certain way, but is it the production of an effect? On the face of it the answer is ‘no’. It is basically a spatial change, and the boulder need not have done anything. Another example he uses is that a boulder can fill a receptacle. That does not seem to be a causal interaction of any kind. So far as I can see, then, the notion of an activity is not to do necessarily with causation, nor even to do with processes since it can just be occupying a space.

Now in the light of these cases I think that it is hard to feel confident that we know what is meant by ‘law-like *activity*’ or ‘kind-characteristic *activity*’.

A further query, though, can be introduced by what seems to me a profound remark by C. B. Martin that I once heard him make. He said something *like* this, ‘Take a case where a ball smashes a window. We say that the ball broke the window. But what did the window *do*? Well, we say that it broke, but it also *slowed down the ball*’. Now, Martin might have meant us to learn a number of things about causation and dispositions from this remark, but what I want here and now to take away from it is that the ball is not just a thing with active capacities, capacities for doing things such as breaking windows, but is also a thing with what we might call *reactive* capacities, such as a capacity to be slowed down in such an encounter. So if you are thinking of what a macroscopic object is, it can be said that it is a locus of *reactivity*, as well as of activity. Take Martin’s case again. When the ball hits and breaks the window it not only slows down since it is slowed down by the window, but the surface of the ball that contacts the window also flattens somewhat. These capacities, which are not in any straightforward sense capacities for activity but are rather capacities for reactivity, are as much amongst the features of that ball governed, as we would say, in a law like way, given what it is, as are the activities it engages in. There is no justification for regarding a thing as more closely tied to its intuitively active dispositions than its intuitively reactive ones.

But continuing in exploring this we might notice a third possible type of case. There are, perhaps, some encounters which occasion no active response, nor any reactive response. Thus, suppose someone drops some water on my hand and it lands and simply runs off my hand. As we might say, I don't do anything to the water, nor does it do anything to me. It is a non-reaction, or perhaps a non-event. However, the outcome of no-reaction is surely as law governed as when there is a reaction or a response. That is, it is part of the nature that I have that nothing (much) happens in such a case. When the outcome is what we might call a null outcome as far as I, the macroscopic object, am concerned, that is as much a reflection of my nature as the other active responses or reactive responses.

So I am suggesting that if we think about the idea of what a MC most fundamentally is, and we reach for its capacities or dispositions at that point as Madden seems to be doing, we should not just attend to so-called active outputs, and also reactive outputs, but also to encounters which produce neither of those outcomes. All of these responses to interaction reflect the nature of the thing, and each is as law governed as the rest. If these comments are fair there is no serious problem for the approach Madden is developing, but in reading it we need not to be misled by what might be a natural way to take talk of activities.

There is another aspect of Madden's treatment of activities that has not so far been mentioned but which strikes me as not straightforward to understand. He says, 'Note that the activities characteristic of a macroscopic kind *K* are macroscopic activities of a whole individual *K* rather than its small parts'.⁴ Now, by activity of a whole individual one might mean something like an activity (or response) in which the whole of the individual *K* is involved. For example, if I roll down a hill then the whole of me travels down that hill. Or if I stand on scales to be weighed then the whole of me affects the scales. We might label this the 'material reading' of 'whole'. On the other hand, there is what we might call the 'linguistic reading' which means activities we report by ascribing a predicate to a term for the entity. Thus, we can say, 'The dog snored'. That is how we report such an occurrence. But that example reveals that the material reading and the linguistic reading can come apart, since it is hard to think that the whole of the dog was somehow involved in the snoring. But there also seem to be examples going the other way. When blood is pumped round your body that seems to involve more or less the whole of you. But we don't say, 'You pumped blood round your body', rather we say that your heart pumped the blood.

Now, if the material reading and the linguistic reading do come apart what can we make of Madden's restriction? It seems to me that he clearly regards as activities relevant to his theory things which cannot be said to involve the whole animal. For example, he cites visualising as an activity, but that is surely done by a part of the animal, and not the whole. In which case he must be relying on the linguistic reading, since we do say, 'The man visualised a so and so'. But this invites the following question: why rule out in our analysis of

what persistence involves activities which are not smaller or less significant or less reflective of our nature and capacities than others that are brought in simply because *we report them differently*? To that question I see no obvious reply.

So far I have tried to bring out that there is something less than clear about the talk of capacities for *activity*, and that the emphasis on activities of the *whole* entity is dubious.

(2) I want next to raise some questions about the talk in the theory of what is *fundamental of types or sorts*. Now, I suspect that many of us talk about what we fundamentally are, or the basic sort of thing that we are, when we discuss these issues. So it is not in any sense distinctive of Madden's treatment. But seeing it very clearly expressed prompts questions that, perhaps, elude us when they really should not.

In his analysis of what an MC 'most fundamentally' is Madden brings in the idea of the general kind it belongs to; he also talk of the 'fundamental kind' the object belongs to. This way of speaking raises two questions. (Perhaps other questions besides these suggest themselves, but I shall respond to these two.) (1) What is meant by 'general kind' and 'fundamental kind'? (2) Why should they be brought into the analysis of persistence of a continuant? Now, in response to (1) we need to be led by Madden's examples. One example he uses is that of a toaster.⁵ He says, 'Take an artefact of the toaster kind'. His main example of a kind is of course 'human animal'. So that is the kind of kind Madden thinks capture what entities *most fundamentally are*.

So given these examples we can ask, why does Madden say that an MC, say me or you, is or are most *fundamentally* a locus of 'law-like activity characteristic of its general kind'?

Well, it is presumably obvious that I have dispositions, which are law-like in their status, but which are not themselves characteristic of the general kind 'human animal'. Thus, I was born with a nature that meant that I have greeny brown eyes – but that specific disposition is not shared by humans as a kind. Other humans have quite different eye colour dispositions. Now, what has not been *explained* is why the dispositions that I share with the kind – human animal – are more fundamental to my nature and persistence than the ones I do not share. From the point of view of law-like dispositions there is nothing more fundamental to the former as compared to the later.

One way to put this is to use a response you sometimes get when you express animalism by saying that we are animals. The reaction is to say, 'But nothing is just an animal; we must be animals of a certain kind'. Now, of course, we should agree with that comment since nothing is *just an animal*. But as a critical response to the claim that we are animals it totally lacks bite. It is like objecting if I tell you that Alex comes from Oxford by saying 'Oh, no one *just* comes from Oxford, he must have come from some part of Oxford'. This is not a valid criticism of the truth of my remark, nor is it a criticism of its usefulness or its informativeness, but the response is nonetheless correct in what it says. So we can also say to someone who says that I am a human animal that no one is just a human animal. You have to be a sort of

human animal. And what that amounts to conceding is that I am not fundamentally a human animal, fundamentally I have to be more than that. Or as we might put it, if we are going to talk of what is fundamental then *there is no obvious reason to stop there*.

So I want to say that if my persistence as a continuant is to be given in terms of preservation of capacities, no reason has been given to single out capacities shared by the chosen kind, human animal, notwithstanding that we can agree that I am a human animal. We should not agree, I suggest, that I am *fundamentally* something with law-like *dispositions shared by the kind* – human animal – and if I am not then we are lacking a reason to analyse continuity of individuals, such as me, in terms of the preservation of those shared capacities.

Leaving that aside for the moment, how does the condition called ‘Persistence’ work? It needs stressing that the condition for persisting is centred on the preservation of capacities, and not on what might be called performances or exercises of capacities. An object may have lots of capacities which it does not exercise, and so focussing on capacity preservation rather than the preservation of activities looks sensible. However, the first example he focusses on is that of a toaster.⁶ His claim is that if we unplug a toaster it retains its capacity to toast bread, and so remains a toaster, but if we break it up and retain only the wire and plug that belonged to it the capacity to toast bread would have been lost and what we have is not a toaster. Both the claims about capacities seem right and so do the claims about what remains in existence in the two cases. But can this be seen as an application of the principle called ‘Persistence’? There are two reasons why it is hard to see it this way. The first is that in the way we think of the persistence of a toaster it is not necessary for it to retain its capacity to toast bread. Toasters break, they undergo accidents and parts fuse together, and yet I can point to my defunct toaster and say ‘this toaster used to make wonderful toast’. The broken toaster is still a toaster. I think that is what we say in such cases. And this possibility seems illustratable by other cases. If the filament in a light bulb breaks it is no longer capable of lighting up a room, but it is a light bulb nonetheless. So, toaster persistence does not fit Persistence.

The second problem is that objects which belong to the same artefactual kind, such as toasters, can be very dissimilar. They can be made of different materials, they function in different ways, and so they are not naturally describable as a sort which are united by law-like characteristics. Each individual toaster will have capacities which are law-governed for it, but they need not be shared by toasters as a kind. This means, it seems, that neither what I called MC nor Persistence applies to artefacts. If you are persuaded by this argument then it is true to say that Madden’s treatment of the link between the persistence of functional artefacts and the retention of capacities for functions is dubious.

However, that can be set aside because for us the question is whether Madden has a plausible approach to animal persistence. It would be quite

natural to view persistence in the animal case as being different from that of artefacts. I want, though, to direct attention to some things about its conception of the animal case which, it seems to me, need thinking about.

Madden combines the general principle which he calls A-persistence with a list of what he describes as 'activities characteristic of human animals'.⁷ This list is eight lines long and contains about fifty activities, and it is also open-ended and clearly envisaged as extendible. It is cleverly thought out and impressive. It includes pointing, walking, running, jumping, tool using, and gossiping, to name but a few. Now, the problem case is my six-month-old granddaughter, who is clearly a human being and has persisted for at least six months but who has no capacity currently to do any of these quite basic things, nor many others on the list, though she has the capacity to do some of the others. There is, it seems to me, a reading of the A-persistence account combined with the (beginnings of the) list which does not fit her persistence. That is to understand the formula as saying that her persistence consists in the preservation of a large number of the listed (an as yet unlisted) capacities. As formulated there is one option to cope with this sort of case. That would be to say that the term 'sufficient' allows for different requirements in different cases. Let us call the final total list of characteristic capacities T. Now, when read with a six-month old in mind the requirement is something like T minus L (six months), where L (six months) contains a large part of T determined in some way by the age of the human in question. And then when she survives the next six months the set to be subtracted from T would be L (twelve months), containing less than L (six months). We would have to understand 'sufficient' along such lines (though we need to add that the notion of 'sufficient' must have more work to do than that). But that cannot be the right way to understand 'sufficient'. The problem based on the present case is that something might go wrong with the developmental processes in the little girl so that she persists but does not acquire the capacities that a typical-year old acquires. That would be a case of persistence but not in accordance with the present reading of the formula. It begins to look as if it is not quite clear how to read A-persistence together with the list in a way which fits the rather sad case envisaged here.

However, this problem is more general than that type of case. The list is meant to be a list of characteristic activities of human animals. It should include what in the language of Putnam was called the stereotype of human animals. That is the shared general picture of human animal attainments that we would expect a normal human to possess. But, of course, it should contain more, since the shared stereotype leaves out what specialist students of human animals would know about but the rest of do not. Parts of it, though, at least should correspond to our shared concept of a human animal. T will contain the evolution of capacities in a typical life of a human animal. In this way the list links with our classification scheme in which human animals are different from, say, gorillas. Using a different way of speaking, this list links to the sortal expression 'human animal'. What, though, does this have to do

with persistence of human animals? The problem is that human persistence as a general thing does not require that lives *take the typical shape*. Lots of human animals do, and all of us could, persist or survive without acquiring the typical set of capacities. In terms of A-persistence the problem is the ‘only if’ component in the analysis. We would agree that if a human animal continuously preserves the capacities for the activities in L it is persisting, although that is not to say anything along the lines of its persistence ‘consists in’ that, but persisting does not *require* any link with the bulk of the content of L.

Perhaps, then, we should regard the list L as not needed in its present form. The basic idea might be expressed as requiring that a human being who has attained a certain level of capacities persists so long as he or she retains a sufficient number of those capacities. This of course leaves what ‘sufficiency’ amounts to in need of further development. However, trying to understand it in this way faces three serious problems.

First, the way of thinking of persistence within this approach is primarily in terms of preservation of (a sufficient number of) certain features. Roughly, for an object to persist things have to remain to a sufficient extent the same. But is there a link between persistence and preservation? One thing to remember is that some animals undergo radical changes in the course of their development. What happens is that they change radically, and gain new capacities and lose old capacities. One hackneyed example is the life of butterflies, which goes through various stages of radical change. Now, in such a case it seems wrong to focus on preservation of pre-existing capacities. Rather there is a radical shift, loss and acquisition, of capacities. Preservation is not what enables us to trace the same butterfly over time.

But, second, even in ordinary developments of human lives, we are not especially looking at preservation. In some cases we are looking at development and acquisition. If we allow that the foetus is a human animal or organism we are not especially interested in whether it preserves its already acquired capacities but in whether it is developing normally. Is its route of acquisition and change normal? Not – has this foetus preserved sufficient of its capacities.

Third, when we trace people around us over time we do not focus on preservation. In cases of traumatic injury the patient persists but preservation of capacities is not required. We accept that there are cases where capacities are radically lost. In normal ageing capacities are lost. No one dealing with us at those stages links the persistence of a human animal with preservation of capacities.

I think that the upshot of these reflections is that there is something out of focus in linking what is involved in human animal persistence to *preservation of capacities*. There is, I want to suggest, also something else questionable in focussing on capacity preservation. A capacity is separate from how it is preserved. Someone can at a time walk, then lose the capacity because of severe leg injury, but then regain the capacity when given an artificial limb. But this means that Madden’s idea of capacity preservation attaches no importance or significance to the grounding of the capacities. Rather, it focusses

on the preservation of capacities. But this runs what might seem to be a risk of allowing that the gradual replacement of organic parts by inorganic parts can amount to human animal persistence since function is preserved but the transformation means nothing of the animal survives. Now, I am not wanting to say that this would be an incorrect verdict. Indeed, I do not see anything problematic in implanting non-animate prosthetic devices to reverse neural failures. However, total replacement even if preserving capacities is surely not obviously enough for an animal to survive.

6 Conclusion

My aim has been to generate some worries about the way that Rory Madden proposes we should think about animal persistence. My worries were occasioned by encountering his account. I have not objected here to his claim that animalism implies that we go with the brain (or cerebrum). I have merely tried to place some question marks against his analysis of persistence.⁸

Notes

- 1 In making that remark I had in mind the total lack of any mention of emotion in Locke's list. But it needs adding that Locke does hold that 'person' is a forensic term and he seems to think that one implication of that is that persons must have something like emotion. So, perhaps, that charge against him is unfair. However, a fundamental issue for people who wish to define the conditions for personhood is what point the term has. And when Locke talks of it being a forensic term he might mean it has to do with responsibility, or he might mean it has to do with what is owed to persons. If the latter then it is not clear we can excogitate a link to emotion.
- 2 For a recent exposition of Wiggins' view see Wiggins (2012).
- 3 Madden (2016), 4.
- 4 Madden (2016), 5.
- 5 Madden (2016), 4.
- 6 Madden (2016), 4.
- 7 Madden (2016), 6.
- 8 I wish to thank Jörg Noller for the invitation to take part in the conference in Munich which resulted in this volume, and also for his patience in waiting for the written version. The conference participants gave helpful responses to my own reflections, but more importantly provided me with many new insights, for which I wish to thank all of them.

Bibliography

- Madden, R. 2016. Human Persistence. *Philosophers' Imprint* 16, no. 17, 1–18.
- Wiggins, D. 2012. Identity, Individuation and Substance. *European Journal of Philosophy* 20, no. 1, 1–25.

4 Personal identity – a process account

Godehard Brüntrup

1 Summary

In this chapter we will pursue the question of personal identity within the framework of process ontology. The central idea of process ontology is the critique of the notion of “substance.” This notion seems, however, indispensable if it is to be really *me*, the identical person, which is surviving through time. It is the very definition of a substance to be that which endures through time.

One main focus will be to establish connections between themes of current analytic metaphysics and traditional Whiteheadian process ontology, ranging from the identity of particulars through time, the metaphysics of time to the reality of abstract objects. Thus, the theory that is sketched here uses in some crucial respects selected Whiteheadian ideas but places them within the debate of contemporary metaphysics. The leading intuition is this: Most process philosophers are realists with regard to processes and idealists with regard to substances.¹ Process ontology assumes events and processes made up of events with a mind-independent unity and identity. Some of these are directly given in experience. Substances, however, are theoretical constructs which are not directly given in experience, they are rather the result of an *abstraction*. It is this central thought of process philosophy that will be exploited here in order to establish an alternative to both endurantism and perdurantism in contemporary analytic metaphysics of personal identity.²

2 The problem of becoming: *metaphysics* between Heraclitus and Parmenides

In his opus magnum “Process and Reality”³ Whitehead presents a surprising interpretation for the biblical verse “Abide with me; fast falls the eventide” (Luke 24:29).

Here the first line expresses the permanence, ‘abide’, ‘me’, and the ‘being’ addressed; and the second line sets these permanences amid the inescapable flux. Here at length we find formulated the complete problem of

metaphysics. Those philosophers who start with the first line have given us the metaphysics of ‘substance’; and those who start with the second line have developed a metaphysics of ‘flux’. But, in truth, the two lines cannot be torn apart in this way.

(PR 209)

The remark “in truth, the two lines cannot be torn apart in this way” expresses in a nutshell the theoretical framework of this chapter. The guiding intuition will be to find middle ground between the extremes of absolute flux and changeless invariance through time. In analytic metaphysics these opposing positions are labeled perdurantism (4d-view) and endurantism (3d-view). According to the 3d-account, concrete enduring particulars like animals have spatial but no temporal parts. They are extended in space but not in time. If you, say, meet a human being you encounter this entire person, not a temporal part of her at this specific point in time. This is the view of classic substance metaphysics. A substance that endures through time undergoes only accidental change. It endures as a numerically identical entity through time. In contrast, perdurance is a continuity of temporal parts in which certain structural similarities are preserved. By adding time as the fourth dimension, the perdurantist claims that concrete particulars have temporal parts. One never encounters an object in its entirety, rather one is in contact with but a time-slice of it. The concrete particular is thus not an enduring substance but a four-dimensional space-time worm, which comprises all temporal stages of this individual (the “worm view”). Or it is claimed that the concrete particular is not even this in space and time extended thing but collapses ultimately into a mere sequence of causally connected stages without assuming any genuine overarching unity (the “stage view”). In both cases the traditional substance view has been abandoned. The idea of substances that are moving through time without being affected in their very essence by the dimension of temporality has been challenged by many contemporary philosophers. Historically, at least since Hume and Locke, there has been a long list of prominent critics of the notion of a substance which cannot adequately be dealt with here. The world view advocated by the sciences, esp. relativistic physics, was in favor of a four-dimensional account (the “block universe”).⁴

3 The internal coherence of a process

We will assume that the human body is a sequence of non-identical physical stages. All there has to hold between these stages is the right kind of relation: immanent causation. This notion was already used by process metaphysician Borden Brown. He was influenced by Hermann Lotze, and his critique of the notion of a substance was tightly integrated with his concept of immanent causation.⁵ The general idea is that a stage S_1 of a given concrete entity E causes a later stage S_2 of E itself. The question that arises immediately is how

E can be construed as persisting entity without assuming a non-changing 3d-substance? Without assuming such a 3d-substance there is no numerical identity through time for concrete entities like human persons. The mere repetition of the relevant properties (stable pattern) together with the right kind of causal connection establishes, however, a weaker form of identity which is often called “genidentity.”⁶ On this view, what we commonly view as one single entity is strictly speaking a temporal series of different entities. Because these entities produce their successor causally while maintaining key properties they can be considered “identical” in the weaker sense of genidentity.

This accords with Whitehead’s view. For him a concrete entity, even a person, consists ultimately of a sequence of psycho-physical events which produce each other causally, while maintaining certain key properties. What we commonly see as an enduring 3d-substance is really a “society” of events ordered serially in time, and thus a process. For Whitehead, an entity which endures in this way through time is characterized by two features: a common element of form and a genetic relatedness that orders the events serially (PR 34). There is a causal inheritance of the defining characteristics in the causal series. The enduring pattern, the form, is not sufficient for individuation. If this were the case, all events that instantiate the same abstract form would be identical.⁷ Of course, one might appeal to the classic Aristotelian idea of special “natural kind” forms whose instantiation generates individuals.⁸ The idea of fixed, clearly and non-vaguely delineated natural kinds is, however, incompatible with the evolutionary world view and with process metaphysics. The basic entities in (Whiteheadian) process ontology are individuated by their unique place in the world as a whole, not by instantiating an individualizing form.

4 Can 3d-objects emerge from 4d-structures?

It might be argued that an individual 3d-object somehow emerges from the repetition of similar events. But even if a stable pattern emerges from the repetition of similar events, it is by no means implied that the stable pattern is a genuine 3d-entity without temporal parts. In process metaphysics, objects enduring in time are, to use a term of Rescher’s, “stability waves in a sea of process”,⁹ patterns of activity that emerge from a base that is in constant flux. It is much more natural to view these stability patterns as processes than to construe them as entities without temporal parts. The thesis that genuine 3d-entities could somehow emerge from a 4d-base of causally connected individual stages or events is not the process view we wish to argue for. It is hard to see how constant flux at the most fundamental level could give rise to real 3d-endurants at a higher ontological level. If an entity that is truly numerically identical through time (no temporal parts) could emerge from a repetitive flux of non-identical entities, we would have to postulate a strong

emergence of substances. A true process ontology can introduce enduring 3d-entities only at the price of incoherence. The same argument can be made against the idea that the higher-level 3d-substances are somehow *constituted* by the underlying series of non-identical events. Constitution theory has been made popular by Baker and others.¹⁰ It builds on the Aristotelian notion that, say, a bronze statue coincides with a heap of bronze by being at the same location in space and time, while not being identical to that clump of bronze. Could an enduring 3d-entity without temporal parts coincide with individual events in a series without being identical to them? That sounds initially promising. But constitution theory is meant to be an alternative to substance dualism. The enduring 3d-entity cannot be an ontologically *independent* substance which then interacts with the underlying events. The substance that is supposed to be constituted by a series of events depends entirely on the constitution base for its very existence. But how could a 3d-substance be generated by a series of events? Thus, we ultimately get back to the idea that enduring 3d-entities somehow “emerge” from an underlying process that is just a series of non-identical events. The intelligibility of this idea has already been questioned.

Emergence and constitution will not explain the existence of substance in a word of flux. We have to find another way. The most basic individuals in process metaphysics are momentary events. If each of these events is causally connected to the following event in the sequence, then the talk of a temporally enduring object can be justified, but only in the sense of genidentity, not in the sense of strict numerical identity. In this context Whitehead often uses “vibration” and “rhythm” as metaphor. An enduring object gains its inner determinations by the rhythmic process of inheriting properties from its predecessors and its own creative novelty (PR 279). This stable rhythmic pattern of its history constitutes the enduring object, which is not a 3d-object but a process. This account is in full accordance with contemporary science. A stable resonance or vibration in a quantum field may constitute what we call a particle. This particle does not exist as a 3d-substance without temporal parts. It is the appropriate connection – the thread of persistence, the stability of the pattern, the genidentity of the underlying events – that justifies the talk of a particle enduring through time.

In what follows, a process ontology of the kind just outlined will have to be assumed. It is the backdrop for the main argument of this chapter. It will have to be shown, however, why the account developed here differs from traditional 4d-views of either the worm or the stage variant. Remember, the aim of our argument is to find middle ground between the “abide with me” and the “fast falls the eventide.” This could neither be achieved with a traditional 4d-view, nor with a traditional 3d-view. Before we can get to this, a few more topics need to be covered at least very briefly because they are central to understanding a process ontology in a broadly Whiteheadian tradition: the metaphysics of time and the metaphysics of abstract objects.

5 Presentism

Presentism is the commonsense and intuitive view of time. Only the present exists, the future does not yet exist, the past does no longer exist. It is well known that this view of time raises some rather difficult philosophical questions, of which the best known is probably the one concerning the difference between the past and the future. The past seems to exist in a certain way, because statements about the past are commonly seen as semantically bivalent (either and only true or false). What makes them true or false if the past does no longer exist? In the current debate the presentist view has been on the defensive due to a criticism David Lewis has formulated in an exemplary fashion.¹¹ Lewis denies that the only intrinsic properties of an object are those it has at the present moment. By intrinsic properties we mean those which an entity has independently of its relations to other objects. Assume that Peter is now blind but could still see ten years earlier. The same person cannot be blind and sighted. The natural solution to this problem is to relate those properties to a point in time. A person can be “seeing-at- t_1 ” and “blind-at- t_2 .” But then we are no longer dealing with intrinsic properties, because we have defined them in relation to a point in time. Thus, for Lewis the only sensible solution is to construe persons as 4d-objects. We can say about the same river that is narrow and that is wide because we are dealing with different parts of the same river. Analogously, we can solve the problem with persons by introducing different person stages. The relativistic view of modern physics with its union of space and time can support this view. In the same way as there is spatial extension in a person there is also temporal extension (and thus temporal parts). The metaphysics of time that is most consistent with this view is the eternalist picture. In the same way as no spatial point has a special status, no point in time (the present) has a special status. The commonsense triad of “past, present and future” (the A-series) is replaced by the duality of “earlier and later” (B-series). A central problem of this picture is that it does not leave room for contingent facts. Any entity in this world has its properties of necessity because there is no open future with alternative possibilities. This view is certainly at odds with our commonsense view. The commonsense view is presentism. Even though the account proposed here is at the most fundamental ontological analysis not a 3d-view, it nevertheless supports a presentist conception of time which is usually associated with a 3d-view. Our position is in fact in some respects similar to a 4d-stage view, where the stages are very brief. However, it differs from the 4d-view by maintaining presentism. This in the contemporary debate somewhat surprising denial of substance and the acceptance of presentism is characteristic of a process ontology of radical becoming.¹² The philosophical motivation for this lies in the desire to take our temporal existence seriously, to be a “serious tenser.” Not only do we regard the future as non-existent, but we experience the past as something which no longer exist, as clearly stated in the expression “thank goodness that’s over!”. This exclamation makes only sense within a presentist framework, as Prior

already noted in his classic paper.¹³ But what about the known objections to presentism? What are the truthmakers of sentences about the past, what makes sentences about the past true or false? This question can easily lead to erroneous ontological claims. For example, if we claim that there once were people that do not exist today. Does this imply that the people of the past still somehow exist? No, it only implies that some people, who do not exist today, *were* existing in the past. To have existed in the past is not the same as having never existed, but it does not imply actual existence. The more difficult question seems to be the one raised by D. Lewis. Can the only intrinsic properties of a person be the ones the person has here and now? Lewis gave a negative answer to this question, but in what follows we will give a positive answer.

6 Endurance of persons in process ontology

What constitutes a human person in this process-ontological account? Is a person defined by mental or by physical properties? In process ontology the mental and the physical are deeply entwined. The mental is the intrinsic nature of the physical. Among others it was Russell who made the argument that physics captures only relational and formal properties of matter that can be expressed mathematically. Everything we know of the intrinsic properties of matter is derived from our experience of mental processes in our consciousness.¹⁴ Without advancing deeply into this metaphysical issue, a point that is salient in the context of this chapter needs to be emphasized: According to process metaphysics, for something to be classified as a concrete particular, it does not suffice that it is manifested as temporally stable configurational pattern in space like, say, a hurricane. A mere aggregate like a cloud formation or a wave is a stable configuration but not an individual. Not any stable process is an individual. A true individual represents the world from a perspective. All metaphysically basic particulars have mental properties, have a unique perspective on the world. This is somewhat reminiscent of Leibniz's monadology. But we do not need to digress into the difficult debate on panpsychism. We just need to focus on persons here. A person has its unity not primarily by the stability of its spatial configuration but by the uniqueness of its perspective on reality as a whole: the first-person perspective. This is a highly relevant feature of a theory personal identity through time in a Whiteheadian process-ontological context. To survive means primarily the persistence of a certain perspective on the world. The question is, however, whether a process ontology has the conceptual resources to explicate the possibility of survival as the persistence of a perspective.

We will define a person as a being that can have a relation to itself as itself, a person has a reflexive self-relation. Other animals have unique mental perspectives on the world as well. Persons are distinguished from most other animals by the fact that they are endowed with a first-person perspective, the ability to distinguish between themselves as seen from the third-person perspective and seen self-referentially from their own perspective. Linguistically

this capability is displayed by the use of pronouns like “I” that the person uses to refer to herself. Persons are able to attribute thoughts to themselves as their own thoughts and reflect on them as such. They thus have self-consciousness, not simply phenomenal consciousness. In process-ontological terms the human person is not a 3d-substance but, diachronically, an ordered series of momentary events, which feature both mental and physical properties. The temporal repetition of these events establishes a person-process. The mental properties allow for the act of self-reference that constitutes the first-person perspective. Synchronically, the person is a hierarchical ordering of higher-level events constituted by lower-level events, where the stream of consciousness of a person is located at the highest level. Below that there are many other levels, from biological cells all the way down to elementary particles.

As outlined above, we construe this process ontology in a presentist framework. Only present events exist actually. They are partly determined by their immediate past, which leaves, before ceasing to exist, a mark on the next event. The structural similarity (common form) and the causal connection (genetic relatedness) of the events allow to speak of a process that endures through time. There is, of course, no numerical identity between the events so connected, only genidentity in the sense given above. The classical notion of a 3d-substance has thus been completely abolished. Within this presentist framework the process of a person cannot be construed as a four-dimensional worm-like entity which is extended through space-time. The process itself is radical becoming. Each personal event grows organically from the past, and is thus a re-enactment of its predecessor, without being fully determined by it, thus allowing for the possibility of creative novelty. In spite of the close genetic relatedness of the events in a personal process, deep down the person seems to become something transient that exists for a moment only, then ceases to exist in the next moment: a process of radical becoming. The concept of a person identical through time, which has dominated Western metaphysics, seems to have been given up entirely by process thinkers. *Panta rhei*, everything is flux: a victory for Heraclitus? In our biblical quote “Abide with me; fast falls the eventide” the second half, which stands for radical becoming, seems thus to be eliminating the first half which stands for permanence. Again, without this commitment to presentism this picture would amount to something similar to the stage-version of the 4d-view. The stages of a person would be momentary in the process of a personal life. In any case, my prior stages would actually be temporal counterparts of me. There is no enduring “I” in any metaphysically robust sense of 3d-endurance. A transtemporal unity of the person could only be conceived by adopting a worm-view.¹⁵ But in the worm-view there is also no “I” that endures through time, each individual time-slice, and thus, each individual moment of first-person experience is different from its predecessor. But the worm-view can at least establish the temporally and spatially extended unity of the entire worm. This requires abandoning presentism,

since the worm exists in a way that attributes no special ontological status to the present. In the final analysis, the process-ontological view seems to lead to a disintegration of the unity of the person. The person is just a well-ordered sequence of events, of which only one actually exists at any moment in time. Can the intuition of personal identity through time be rescued in this metaphysical framework?

7 Abstraction and permanence

Whitehead famously remarked that “to be an abstraction does not mean that an entity is nothing. It merely means that its existence is only a factor of a more concrete element of nature.”¹⁶

With that in mind, let us return briefly to Whitehead’s idea that it is a common form of serially ordered events that allows for the recognition of an enduring object. Whitehead writes that the form is a complex eternal object (PR 34). Eternal objects in Whitehead’s terminology are abstract entities. This notion can easily be misunderstood. Whiteheadian “forms” are not Aristotelian forms. They do not cut out 3d-substances. If eternal objects have structuring “impact” in the world by some kind of formal causation, then we have arrived at a Hylomorphic Aristotelian view. A timeless *forma substantialis* configures matter in such a way that an enduring 3d-object is constituted. If this substantial form were doing all or most of the metaphysical work in the individuation of a concrete particular, a classical substance ontology would be re-established. That is incompatible with process metaphysics which means to overcome the notion of substance. In what follows the ontological status of those abstract objects or forms will thus be “downgraded” to a non-primary or derivative status. The ontology primacy remains with the actual events. For this purpose, it is essential to clarify what is meant by “abstract.” Typically, abstract entities can be realized at different places and at different times. The abstract form “triangular” can be realized at different times and also simultaneously at different places. Concrete particulars, which we might call “continuants” to avoid the loaded notion “substance,” can also exist at different points in time but not simultaneously at different points in space. One and the same person can exist in 2005 and 2010 but not simultaneously at two locations in space. Continuants are thus similar to universals with respect to multiple temporal and non-simultaneous spatial realizability.¹⁷ Following a tradition that reaches from Neo-Platonism to Leibniz, Whitehead assumes that all basic entities are concrete entities. Abstract entities are not self-grounded; their existence depends on the activity of minds in concrete entities. This position that is located between realism and nominalism is often called “conceptualism.” Abstract objects are *entia rationis*, they are mind-dependent. If, in the tradition of Neo-Platonism, it is the divine mind that secures the existence of abstract objects, one might as well speak of a realist position. If confined to the human mind, it is rather a

form of anti-realism. As noted earlier, continuants are indeed similar to universals by being able to be instantiated at different times. If abstract entities are conceived as entities that are dependent on thinking concrete entities for their very existence, then a middle ground between a pure 4d- and a pure 3d-view is indeed possible. At this point one can refer to the above-mentioned thesis by Nicholas Rescher: Process philosophers are realist with regard to processes and idealist with regard to substances.¹⁸ In the following we will build on this basic intuition.

In the process-ontological account of presented here, only momentary events that are related by genidentity are in the right relation to form a chain of persistence, or a “society” as Whitehead would put it. Two conditions must be met: causal dependence and common form. We can then speak of “immanent causation.” What is relevant here is the common form as a multiply realizable abstract entity. Entities connected by the relation of genidentity share this common form. The abstract entities remain unchanged through the unfolding process. They are invariants of the genidentity relations. The salient point is this: Abstract entities are assumed to be ontologically mind-dependent. The analysis of the mental process of abstraction will tell us more about the exact nature of these entities. The classic view of abstraction as a filtering of common characteristics in a multitude of similar cases was replaced in more recent philosophy by a conception that can be traced back to Frege. Frege noted in the *Grundlagen der Arithmetik* that many of the singular terms referring to abstract entities derive from functional expressions. We speak of the “*number of objects*” and the “*direction of objects*.” “Number of...” and “direction of...” are incomplete (“*ungesättigte*”) functional expressions.¹⁹ The genuine discovery by Frege was that typically for functional expressions that single out abstract objects, there are equations of the following structure:

$$f(a) = f(b), \text{ iff } a R b,$$

where R is an equivalence relation. To use Frege’s example:

The direction of a = the direction of b , iff a is parallel to b .

The number of F s = the number of G s, iff there are just as many F s as G s.²⁰

The meaning of “number” is determined by the equivalence relation “just as many” or “equinumerous.” Frege hinted at this theory of abstraction, only recently has it been more fully developed by Crispin Wright and Bob Hale.²¹ Because equivalence relations are reflexive, symmetrical, and transitive, they can be used to introduce continuants that preserve their identity through time. We are dealing with a theory of abstraction, however. Continuants become abstracta in a very specific sense. Peter Simons makes use of this Fregean intuition when he introduces continuants in his ontology, which is basically a 4d-account that does not recognize 3d-substances.²² It might be useful to clarify the basic idea a little more. Take a number of objects over which an equivalence relation has been defined, say, an equivalence relation

with regard to their mass. It may be called “equi-massive.” Assuming the analysis by Frege one can engage in the process of abstraction:

a is equi-massive to b

can be conceptually transformed into

the mass of $a =$ the mass of b .

Thus, the abstract idea of mass has been introduced, and the term “mass” refers to it. Now let’s apply this procedure in the context of this chapter. Let’s call entities that are endowed with a perspective on the world “perspectival.” We define an equivalence relation over the set of entities that are perspectival:

a is equi-perspectival to b

can be conceptually transformed into:

the perspective of $a =$ the perspective of b .

This can be done for first-person perspectives as well:

a is first-person-equi-perspectival to b

can be conceptually transformed into:

the first person perspective of $a =$ the first person perspective of b .

So far we have been working within the framework of the theory of abstraction developed by Frege, Hale, and Wright. In the context of personal identity through time this account needs to be expanded because a and b exist at different points in time, thus we get:

a at t_1 is first-person-equi-perspectival to b at t_2

This can be conceptually transformed into:

the first person perspective of a at $t_1 =$ the first person perspective of b at t_2 .

The first-person perspective is the identity criterion for persons. Two persons are identical if they have identical perspective on the world such that they are able to refer to themselves as “I.” One arrives thus finally at:

the person at $t_1 =$ the person at t_2 .

8 Personal identity

What has been developed above is – in a nutshell – a theory of personal identity. The most striking feature is that, according to this account, the concept person refers to an abstract object that was introduced by an equivalence relation. The question that immediately comes to mind is this: How can two events featuring a perspective considered to be equi-perspectival? Identity of perspective must not be used to ascribe equi-perspectivally because that would entail a vicious circle. The abstract notion of a perspective ought to be derived. The equivalence relations must thus be introduced independently of the abstract entities that will be derived by these equivalence relations. This is where the process ontology developed above helps fill the gaps. A suitable equivalence relation stands in need of certain stable patterns and the appropriate causal connection. Because these two elements are sufficient for establishing genidentity they are also sufficient to establish equivalence relations. The “thread of persistence” (P. Simons) between the events in a temporal sequence is genidentity. Genidentity rests, as already pointed out by Whitehead (PR 34), on the appropriate causal connection and a common element of form. The thread of immanent causation thus established allows for a multitude of momentary events to be joined into an enduring, stable process. But the process is not yet a 3d-object, it is a stable rhythmic repetition of similar events. The 3d-object, according to our key claim, does not exist independently of mental abstraction working with equivalence relations. It is an *ens rationis*, something that exists through thinking. A thinking mind can make use of the equivalence relations that are based on causal relations and common forms: a mental process that produces via abstraction the notion of a 3d-continuant. Continuants are therefore abstract objects, which can be realized at several points in time and non-simultaneously at several spatial locations. They are well-founded in the reality of the appropriately related events. *It is thus by our mental activity that we introduce stability into the world that withstands the eroding power of the Heraclitean flux.* This account does indeed imply a sort of Berkeleyan idealism with regard to 3d-substances. Their being is partially a “being conceived” as continuant. To the question that was raised against Berkeley as to what happens with them if nobody is thinking of them, the theists among the process metaphysicians might well answer just like Berkeley: God secures their existence with his omnipresent mind. One could, however, bite the anti-realist bullet and grant that there are no 3d-substances in the world that exists independently of the human mind. The introduction of mind-independent substances might then be conceived as the “original sin” of Western metaphysics.²³

9 Taking stock

Rescher’s claim that process philosophers tend to be idealists with regard to substances was spelled out by drawing on a theory of abstraction originally

introduced by Frege. The result was an abstractionist view of 3d-continuants. Continuants without temporal parts are abstractions that we introduce in order to structure our physical environment, and possibly even more so our social environment. Their introduction is not arbitrary, rather it is founded in the causal relations among events that generate stable patterns through uniform repetition; the underlying relational structure is that of genidentity and immanent causation. Sentences about 3d-substances that are ontologically committed to the existence of continuants are not strictly speaking false. In the same way the talk about centers of gravity is not false in physics, even though centers of gravity are abstractions that do not exist strictly speaking. We are arguing for an abstractionist view of persons that is founded in concrete and real causal patterns. The continuant is not, however, simply identical to a sequence of momentary events that constitute its life. That would amount to a pure 4d-view. It seems that the view suggested here can prove more fruitful for understanding our social practices than a 4d-view. The 4d-view drops the notoriously difficult notion of a substance entirely and settles instead for time-slices in worms or stages that are connected by appropriate relations. The account suggested here can establish ontologically well-founded 3d-continuants which are introduced by abstraction. The 3d-view is saved, but not in its original metaphysical form. Enduring substances are ontologically dependent on minds and as such differ from classically construed 3d-substances. One advantage of this view is that it does not require dropping our commonsense 3d-metaphysics. It is a practical necessity to introduce time-invariant fixed points in our commonsense world view. The classical notion of a substance serves this purpose. The continuants of the abstractionist view advocated here are, however, fully present at each passing moment. They serve the same function as the classical concept of a substance. In contrast, a revisionist metaphysics implying that we cannot encounter entire persons as such but only time-slices of persons or person-stages is hardly sustainable in a life lived according to customary social standards.

The view proposed here can be further elucidated when contemplating the possibility of fission. What happens if two or more spatially separated personal events are connected in the right way to a sequence of earlier personal events? This is a deep puzzle that cannot easily be solved within a metaphysical account that assumes a 4d-view at the most basic level. In the 4d-view, seen from the two persons that exist after the fission, two continuants overlap in the past. Two persons coincide spatiotemporally. That is a problematic notion. Here the process-ontological view does better than the traditional 4d-view because the continuants are *entia rationis* and thus mind-dependent. From each personal point of view, there is thus indeed only one abstracted person in the past, and there is no deeper mind-independent level of overlap. There is just one sequence of events related by genidentity. Two competing abstractions can co-exist without causing problems at the most basic metaphysical level of concrete events. Deep down there are only individual events, none of them numerically identical with any other in the sequence. Since

the enduring persons are of our making, *entia rationis*, it is up to us to settle the social consequences of such a case were it ever to arise. It is doubtful, however, that if a single earlier event in nature has two different simultaneous successors that both are connected by genidentity to the earlier event. A more natural way might be to view this fission as the beginning of two new causally connected processes.

Interestingly, in an ontology of radical becoming, death loses some of its uniqueness. Dying is happening constantly in the transition from one momentary event to the next. The fact that we are not worried by this fact is grounded in the experience that in natural life each experienced moment is expected to have an immediate successor which is connected to the earlier one in the appropriate way to establish genidentity. As we saw earlier, the concept of genidentity is associated with the concept of immanent causation. The general idea was that a stage S_1 of a concrete particular E generates or brings about a later stage S_2 of E itself. Above, we have applied this idea to the problem of personal identity: A human person P which exists at, say, time t_5 is genidentical to a human person which existed at t_1 if the temporal stages that lead to t_5 are immanently causally connected to the earlier temporal stage of P at t_1 . This is the case in our daily experience. The events in the flux of consciousness experiences have the causal power to bring about their immediate successors. The stream of consciousness, our phenomenal experience from a first-person perspective, is such a chain of serially ordered events. This is a process that, under normal circumstances, self-perpetuates during the span of a human natural life. From this experience of a self-perpetuating process, we abstract the notion of a numerically identical person enduring through time. Persons as 3d-objects enduring through time are mind-dependent. They are abstractions. That is not denying their reality. The quote by Whitehead that was given above makes perfect sense in the case of personal identity: “To be an abstraction does not mean that an entity is nothing. It merely means that its existence is only a factor of a more concrete element of nature.”²⁴

Notes

- 1 Cf. Rescher (1996), 58f.
- 2 An early ancestor of this chapter appeared in: Gasser (2010), 67–85. There are, however, substantive changes in form and content.
- 3 Whitehead (1978).
- 4 The idea is already quite clearly developed in Minkowski (1909), 75: “Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.”
- 5 Bowne (1882):

A change between things must depend on a change in things. Now when we remember that the only reason for positing things is to provide some ground for activity and change, it is plain that the changeless core is of no use and must be dropped as both useless and unprovable.

- “Interaction cannot be conceived as a transitive causality playing between things; it is rather an immanent causality in a fundamental unitary being” (83). For the most recent developments cf. Zimmerman (1997).
- 6 The concept “genidentity” was originally developed in Levin (1922).
- 7 I am not pursuing the interesting idea of individual essences here. It seems hard to square with the basic intuitions of process philosophy where the individual event is not fully determined by any pre-existing essence.
- 8 Cf. Loux (2006).
- 9 Rescher (1996), 53.
- 10 Baker (2007).
- 11 Lewis (1986), 202f.
- 12 In the current debate, Galen Strawson has advocated a similar view, and he then claims that the “Persistence Belief is not experientially natural” (Strawson 2009, 221).
- 13 Prior (1959).
- 14 Russell (1927), 270, 402.
- 15 Cf. Sider (2003).
- 16 Whitehead (1920), 171.
- 17 Cf. Simons and Melia (2000).
- 18 Cf. Rescher (1996), 58f.
- 19 Frege (1884), §§62ff.
- 20 Cf. the excellent presentation of this topic in: <http://plato.stanford.edu/entries/abstract-objects/>.
- 21 Wright (1983); Hale (1987); Hale and Wright (2009).
- 22 Simons (2008).
- 23 This view has been advanced recently by Puntel (2008).
- 24 Whitehead (1920), 171.

Bibliography

- Baker, Lynne (2007). *The Metaphysics of Everyday Life: An Essay in Practical Realism*. Cambridge: Cambridge University Press.
- Bowen, Borden (1882). *Metaphysics: A Study in First Principles*. New York: Harper&Brothers.
- Frege, Gottlob (1884). *Die Grundlagen der Arithmetik. Eine logisch mathematische Untersuchung über den Begriff der Zahl*. Breslau: Wilhelm Koebner.
- Gasser, Georg (ed.) (2010). *Personal Identity and Resurrection*. London: Routledge.
- Hale, Bob (1987). *Abstract Objects*. Oxford: Basil Blackwell.
- Hale, Bob, and Crispin Wright (2009). “The Meta-Ontology of Abstraction.” In: D. Chalmers, D. Manley & R. Wassermann (eds.), *Metametaphysics. New Essays on the Foundations of Ontology*. Oxford: Clarendon, 178–212.
- Levin, Kurt (1922). *Der Begriff der Genese in Physik, Biologie und Entwicklungsgeschichte: Eine Untersuchung zur vergleichenden Wissenschaftslehre*. Berlin: Springer.
- Lewis, David (1986). *On the Plurality of Worlds*. Oxford: Blackwell.
- Loux, Michael (2006). *Metaphysics. A Contemporary Introduction* (3rd edition). New York: Routledge.
- Minkowski, Hermann (1909). “Raum und Zeit,” 80. *Versammlung Deutscher Naturforscher* (Köln, 1908). *Physikalische Zeitschrift* 10, 75–88.
- Prior, Arthur N. (1959). “Thank Goodness That’s Over.” *Philosophy* 34, 12–17.
- Puntel, Lorenz B. (2008). *Structure and Being*. University Park: Penn State University Press.

- Rescher, Nicholas (1996). *Process Metaphysics. An Introduction to Process Philosophy*. New York: Suny Press.
- Russell, Bertrand (1927). *The Analysis of Matter*. London: Routledge.
- Sider, Theodore (2003). *Four-Dimensionalism. An Ontology of Persistence and Time*. Oxford: Clarendon.
- Simons, Peter (2008). "The Thread of Persistence." In: Christian Kanzian (ed.), *Persistence*. Heusenstamm: Ontos, 165–184.
- Simons, Peter, and J. Melia (2000). "Continuants and Occurents." *Proceedings of the Aristotelian Society, Supplementary Volumes* 74, 59–92.
- Strawson, Galen (2009). *Selves: An Essay in Revisionary Metaphysics*. Oxford: Oxford University Press.
- Whitehead, Alfred N. (1920): *The Concept of Nature*. Cambridge: Cambridge University Press.
- Whitehead, Alfred N. (1978): *Process and Reality* (PR). Corrected Edition. New York: The Free Press.
- Wright, Crispin (1983). *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.
- Zimmerman, Dean (1997). "Immanent Causation." *Noûs*, Vol. 31, *Supplement: Philosophical Perspectives*, 11, *Mind, Causation, and World*, 433–471.

5 What are persons?

Attempt at a normatively neutral ontological account

Thomas Buchheim

Beginning with Robert Spaemann's (and Romano Guardini's) *dictum* that persons only exist in the plural,¹ and also guided by the assumption that living individuals, insofar as they are persons, are persons *by* nature, that is born persons, but not *in virtue of* their nature, I arrive at four pivotal propositions (P1 to P4), each of which – not only according to my opinion, but following widespread intuitions and, as I think, not an uncommon usage of the word – is necessary, and which together could, in my view, count as even sufficient for calling something a person.²

It is my aim to show that although persons can each be *identified* with a living individual (or, if you want, an animal of higher kind) it is rather their common “form of life” in an assembly of many of them, which qualifies them all as *persons*. A *form* that is realized by the biographic conduct of most, but not necessarily *all* associates in the ensemble. This kind of *form* that is not instantiated as a feature of the individual but rather of the assembly as a whole – although not all members of the association need to share its specific requirements of implementation – is in this regard similar to life forms in the sense of Michael Thompson's “natural historical judgments”.³ Except that the respective judgment like “Man lives the life of a person” is, according to my thesis, just not a *natural* historic judgment, but at best a *vivento*-historical judgment. And I also do not consider the depicted personal life form from primarily a logical point of view as a special kind of *judgment*, but as an ontologically real form⁴ that builds on living associations under certain circumstances that I want to specify in more detail in what follows.

Before beginning, I want to say something about how to understand the “plural” that is necessary for the existence of persons according to my thesis (which is inspired by Spaemann and Guardini).

I take this plurality of persons as a conceptual but not a factual requirement. This means that even if there did not exist, by actual circumstances, more than only one natural living individual of personal life-character, it would still remain a person. To be a *person* requires *conceptually* (not factually) to be one among a multiplicity of them, although the relationship between the concept of a person and its multiplicity is neither essential (like, for example, the concept of a natural number) nor a merely accidental one (like,

for example, the concept of a twin). *What* is a number in general could not be conceptualized without being a particular one out of many, whereas *what* a twin is could very well be understood without the twin. Being a twin is a mere accident for living individuals like twins. This is, however, not the case in terms of persons: Neither is, what any single person is, impossible without a plurality of persons; nor is it negligible for the living of each to exist in the plural. Aristotle would have described the conceptual entanglement of “person” with the plurality of cases as an accident *kath’ hautō*: It belongs to a person *per se* to exist among a plurality of cases, like a planet is *per se* one of many orbs in a solar system. But unlike the planet, a person would remain what she is as a person, even if the pluralistic system ceased to exist some time. Whereas the planet would not be a planet anymore if one removed it from the solar system it belongs to. We shall immediately see why this difference occurs when we get into the explanation of the four pivotal propositions.

I want to present them at the beginning as a kind of overview, in order to explain each of them progressively on the basis of the preceding one:

- (P1) The basis of the existence of persons is the tradition or communication of a *common heritage of life* to a *plurality* of living individuals of accordingly similar nature (=association by filiation).
- (P2) The persons’ common heritage of life is ethologically highly complex (a non-primitive life): This means that we must distinguish with regard to all members of the association by filiation between the *biography* of life, which turns out in highly multifarious ways for different individuals, and the naturally or biologically *pre-given burden of life*, which permanently determines the life of all in a similar way (=differentness of life-biography of variant cases and commonly shared biology of life).
- (P3) Because of (P1) and (P2) there exists a *universal isotopy* of all members of an association-by-filiation – in the sense of a general *perception of the twofold identifiability* of each abstract-arbitrary member of the association, which manifests itself in all biographies (be it actively or passively): (a) Identification by its relationally determined, but nonetheless in principle *interchangeably occupied place* within the respective association; (b) identification by its descriptive (naturally inherited and biographically acquired) properties (=isotopy by means of perception of the twofold identifiability of each member).
- (P4) As a manifest consequence of (P3), the highly variant biographies of the members of an isotopic life-association bear throughout a significant *uniformity* in terms of a situative pattern of *representation* in the behavioral structure of most of the members: A pattern according to which each member that is *able to represent herself* has to *co-represent* the other members of her field of awareness (=biographical pattern of representation).

1 Whatever is a person exists by filiation of life into a multiplicity of living individuals

(P1) *Association by filiation.* The basis of the existence of persons is the tradition or *communication of a common heritage* of life to a *plurality* of living individuals of accordingly the same nature.

By “basis of existence” I mean not only a necessary condition in the common philosophical sense of the word, but more specifically such a condition insofar as it is also a *constituent* of what is conditioned by it and therefore belongs to it. According to this first pivotal proposition it is clear that nothing would be a person which was not alive, and likewise nothing which, as a living being, existed on its own, that is which would be the solitarily sole instance of its animation (*Lebendigkeit*) as long as it exists. Rather, life must be transferred or communicated in a *plurality* of cases of its own instantiation. To speak with Aristotle: The respective life must have a *natural essence* – *physis* as *eidos* – that develops *synonymous cases*, not only be a continued growth without individuating self-limitation like, for example, a tumor.

It is not important for the respective living condition of persons in terms of (P1) whether life is organic, naturally or technically produced or even whether it operates on the basis of matter at all. Rather, I follow John Locke, who – like Aristotle did before – argued that the condition of individuality of living identity does not result from the matter on which life operates, but from the individual closure of life in each instance of its occurrence. I quote a passage from Locke:

animal Identity is preserved in Identity of Life “(II 27, §12 [p. 337]” the same Animal, as we have observed, is the same continued Life communicated to different Particles of Matter, as they happen successively to be united to that organiz’d living Body.

(II 27, §8 [p. 332 f.]

According to the position, I want to defend, persons are always *individual* lives in the sense of Locke’s definition, but – and in this regard contrary to Locke – something is the same person as long as its numerically individuated life continues, which corresponds most likely to the position of *animalism*. For Locke argued that it is not the *same life* which, besides the living individual, accounts for the same person, but rather the same consciousness. According to Locke, even *different* living beings could, in principle, be the same person if they only shared the same consciousness – whatever Locke thought consciousness would be:

we must consider what *Person* stands for; which, I think, is a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places.

(ibid. 335)

In this I do *not* want to follow Locke. Rather, it is essential for my following propositions that the identity and the existence of a person does not rely on the question of her potential or actual self-consciousness. In my opinion there can be *persons* that do not have self-consciousness at all. Furthermore, I think that it is not helpful to assure that a living organism which has at least potentially self-consciousness or a “first-person-perspective” be *thereby* a person although this is a common argument in the contemporary debate of the matter. For first of all it is very unclear under which actual circumstances something or someone has *potentially* self-consciousness or a first-person perspective, and under which not. And because of this it is, second, difficult to decide by means of clear criteria whether an individual – if potentially self-consciousness is granted – is or is not already actually a person or likewise only *potentially*. Here again I agree to the radical position of Robert Spaemann’s, who has categorically argued that “[t]here are no potential persons”.⁵

In what I have tried to explain it remains open what it means that life as such can have individual closure and therefore numerical identity. I presuppose this as acceptable and assumed wherever it would concern persons. Also it remains open what *kind* of individually self-contained life can justify that it is not only the life of brutes – or in Locke’s words – “animal life”, but the life of *persons*. For, of course, it is obvious that not all individually self-contained life which is inherited or communicated in a plurality of instances is already personal life, and that the members of such an inherited association by filiation are already *persons*.

The following three pivotal propositions shall answer the second question: What kind of communicative association of the same heritage of life for a plurality of living individuals provides the characteristics of personality? (Regardless of whether artificial or biological ways of communication are chosen and whether it is a materially bound life or not.)

Before I proceed to the next proposition, I would like to add something more to the first one, which I consider to be typical for persons, which is why I think that it is necessary that persons are *living* beings. For whatever is a person, if the first proposition is true, it can only act *due to* and *within the limits* of its own being alive. The basis and the limits of all its life-manifestations or activities necessarily belong to the *same* individual case of life, like the individual living activity or action which is produced on this background of being alive. This means: For each activity or action of a person, her living nature is an *associated condition* which she cannot preempt by her activity or action.

This holds for persons in a general way. For persons of all sorts it would therefore be conceptually important that their life partly manifests traits not to their own disposition and partly is biographically disposable.

2 The difference between biographical and biological (naturally given) life

From the explanation of the first pivotal proposition we can see what it means that whatever is a person is a person *by* nature in a given and unavailable way.

But not – and this is expressed by the second proposition – also *thanks to* this nature which is given to her, but rather thanks to what persons make of their life, if they have it at all. I want to call the latter, which individually and variably disposes a living being thanks to its aliveness, its *biographical* life. Whereas the former be called their *biological* life which, as a mere being alive, is given to the person by nature (or by some technical means which is impossible to preempt for any thereby existing person).

From what we know as varieties of life on earth, it seems to be obvious that this difference between biographical and biological life is not expressed in all kinds, but at least in many of them, even apart from the human kind. According to my second thesis, we can only consider such assemblies of inherited aliveness as possibly consisting of persons if their individual *biographical* lives are clearly distinct and situationally highly variable compared with a merely biological life or being alive. Therefore, living beings such as sardines or ants would not qualify for persons according to P2, but living beings of high biographical significance, like, for example, dolphins or primates exhibit, which are ethologically highly complex living beings. With reference to the variety of cases within an association of bearers of a similar heritage of life, it follows the second pivotal proposition:

(P2) *Differentness of life-biography and commonly shared biology of life.* The persons' common heritage of life is ethologically highly complex (a non-primitive life): This means that we must distinguish with regard to all members of the association by filiation between the *biography* of life, which turns out in highly multifarious ways for different individuals, and the naturally or biologically *pre-given burden of life*, which permanently determines the life of all in a similar way.

The most important point that follows from this distinction for the essence of persons is the one that I have stressed at the very beginning, namely that persons are, what they are, not *thanks to* their nature or *in virtue of* their being biologically alive, but only thanks to their biographical, although highly diverse life, which notwithstanding the varieties are alike, as we shall see, in certain *formal* regards. That they are *persons*, and not non-personal living beings, is due to their biographical life, what means that they are not actually *persons* simply *thanks to* their biological nature.

This means furthermore: The concept "person" is according to my thesis not a natural concept and does not designate a *species* or *natural kind*. By contrast, the membership of an association by filiation of all empirically accessible persons implies, according to my *first* proposition, that they always belong to a species of a natural kind, insofar as they are not artificially created persons. But even if they were artificially created, they would be several ones of the *same* life-heritage and therefore, if not exemplars of a *natural kind*, after all exemplars of an *artificial kind*, which is prerequisite to each of their individual lives.

So to be a person is, according to P2, in any case a *biographical achievement*, not a biological matter of fact. I speak of an achievement, if the result of a behavior does not occur on its own, like, for example, perspirations, but

has the character of a performance: A performance, however, which is not the objective or purpose of the respective behavior, but goes along with it involuntarily. This is similar, e.g., to the fact that knowing to orient oneself in an area goes along with visiting a place frequently. It is specific to the case of being a person as a biographical achievement that the respective involuntary performance goes along with *any* variant biographical conduct, as we shall see.

Since it is a certain biographical achievement in this highly general sense, it is by no means essentially the matter of one *particular* species to bring about persons, whereas all other species would be left with nothing. The charge of “speciesism”, that is the unjustified privileging of members of the *human* species as persons against all others, is from the outset prevented by the second proposition which I hold. Hence, it does not depend on our species, and it is not a case of arbitrary self-preference or “speciesism”, that we humans take ourselves to be persons, but exclude until today all other species that are known to us from it. Rather, we as human beings have, so to speak, earned this privilege over all other natural species by means of our biographical way of life for reasons that still need to be made explicit.

As I have already suggested, the shared biological life within an association by filiation provides (unlike the biographically highly variable life of each single member) something like a *common burden* or condition for the biographical life of each individual. This imposes upon all members of the respective association a certain inaccessible *equality* of their life and their staying-alive with others which cannot be escaped. Thereby certain corridors are opened to the variability of the biographical life of each member of the association, whose borders are not sharply determined but cannot, as zones of tolerance, be biographically transcended by preserving their life. Therefore, personal life, wherever it occurs, is biographically inaccessibly equivalent to other lives under a similar burden at many other positions within the association. This leads to the third pivotal proposition that I want to argue for:

3 The biographically realized isotopy of personal life

According to the next step of my argument, the naturally grounded equality of the biological life of many consistently manifests itself as a basic situation of each individual also in the *biographical* life of all. This is expressed by proposition (P3), which is, for reasons that I will explain soon, complicated and which gains a specific *personal*, not only natural or biological profile when being split into two branches or aspects.

(P3) *Isotopy by virtue of the twofold identifiability of each member.* Because of (P1) and (P2) there exists a *universal isotopy* of all members of an association-by-filiation – in the sense of a general *perception of the twofold identifiability* of each abstract-arbitrary member of the association, which manifests itself in all biographies (be it actively or passively): (a) Identification by its relationally

determined, but nonetheless in principle *interchangeably occupied place* within the respective association; (b) identification by its descriptive (naturally inherited and biographically acquired) properties.

The word “isotopy”, as I take it, is used in literary studies and means something like “the same topic or *topos* at many places in a text”. In this way, according to the first two propositions, the biographical life-situation of each person is necessarily constituted, no matter to which species they belong. A biographical life-situation is not to be identified with the subjective consciousness that someone has of such a life-situation. The isotopy reaches further than only to these members of the filiative association that are conscious of their situation.

How a certain biography goes in a singular case, which biographically individually variant life is being lived does not only depend on the subjectively made dispositions and behaviors that an individual manifests. Rather, it depends on the kind of objective or coexistential situation in which a life proceeds. For this reason, *all* members of the filiative association are biographically concerned by the existing isotopy, and not only those which perceive this subjectively.

The typical and, according to my thesis, purely *formal* characteristic of the prevalent isotopy that is independent from the special nature of the respective species as well as from the biographic variability of life consists in the fact that each equal member of an association by filiation can be identified in two independent ways: *Either* the identification (a) by means of its position or place within the isotopic association of biographic coexistence, *or* the identification (b) by the descriptive properties that it has due to its natural quality and because of its biographical acquirements. Whereas the identification (b) by descriptive properties does not require any special explanation, the identification (a) by the place within the isotopic association has an internal complexity which one could call a “biographical situation”. For, on the one hand, it is always composed of the place of origin in the course of the filiation or tradition of the life-heritage (which in the case of us humans is fixed by the *name* and often by a birth certificate), and, on the other hand, of a biographically actually occupied place in relation to other isotopic members of the life-association, whose *trace* always biographically reaches back to the first mentioned place of origin.

To occupy such a place as a biographical situation within an isotopic life-association is indeed the case in non-personal life-associations as well and insofar as generally independent from the descriptive characteristics of the respective species. It is, however, *not* equally independent from the nature of the species (although not necessarily committed to only one of them) that its exemplars normally and to a high degree *first* detect the disparity of those *two* ways of identification for members of the association and, as it were, biographically realize it in their behavior; this goes along with a sense for the principally possible *interchangeability* of a descriptively determined setting at one and the same place, which exists in all circumstances of their life.

Second, they are able to *abstractly* perceive that each *individual* (no matter which) that would hold any of the places that are in principle all neutral regarding their actual occupation would be an associated member of this life-association. Since the actual occupation of the place is, however, an identification that is independent from the descriptively identifiable properties, this abstract perception or idea carries a likewise abstract *extensibility* of the association with it that is independent from the specific descriptive properties of the members.

Because of this subliminally perceived *abstract extensibility* it is by no means necessary that each member of a *personal* life-association must, in occupying an isotopic place, also share the same affiliation of species. This is the case although the conditions for the isotopy of occupied places are objectively met by the mere factual belonging to a certain species (for example, in the case of humans). For only factual traceability to an original place within *some* filiative association of personal character in general is necessary, which could, in principle, be an association of *another* kind of living beings. If, for example, extraterrestrial (or artificially created) living beings entered the stage that belonged, due to their own life-heritage, to a filiation of appropriate sort corresponding to the formal requirements of P1-P4, then also these beings would automatically occupy *isotopic* places within *our* human life-association in the sense of P3 (without having to ask for permission). How we would *deal* with such factual extensions of the universal isotopy according to P3 is another question, of course a primarily ethical-normative question, to which I do not have to say more from an ontological point of view. Thereby, however, it becomes clear that no descriptive species-properties that do not belong at the same time to the *form* of personal life are involved in the identification in terms of (a) – namely the occupation of one of the isotopic places in the actual biographic situation.

In his book *Persons*, Robert Spaemann had called this presented isotopy a “relational system”, of which persons as persons would have to be part of,⁶ or, alternatively, a “personal space”,⁷ in which the biographical life of persons would basically be situated. The two biographical characteristics that find expression in the life of all members of the relational system are, as I said, the formal and abstract traits of *interchangeability* (of bearers of properties) at the same place that each person occupies, and the principally unfinished and open *extensibility* of the relational system.

4 The pattern of representation and its ethical-normative neutrality

In doing so, we are approaching the fourth and last pivotal proposition on the general characterization of a *form* of the life of persons. This proposition tries to present a certain biographical pattern in the behavior of placeholders of an isotopic relational system that is sufficiently general to make a claim on *all* biographies within the respective system, and sufficiently specific in order to

hold for a formal feature of associations of *personal* coexistence, independently from any special culture or cultural activity, but also regardless of a possible difference of the species-affiliation of its members. This biographical pattern takes into account the isotopy, which I have described in (P3), in various ways, but it is, contrary to it, not universal, that is it does not apply to all members of an isotopic life-association without exception at any time. Instead, it forms a standard for behavior for the normal case of biographical life *in all variations*, which *attunes automatically*.

(P4) *Biographical pattern of representation*. As a manifest consequence of (P3), the highly variant biographies of the members of an isotopic life-association bear throughout a significant *uniformity* in terms of a situative pattern of *representation* in the behavioral structure of most of the members: A pattern, according to which each member that is *able to represent herself* has to *co-represent* the other members of her field of awareness.

I want to understand the concepts “representation” or “substitution” in orientation of the basic operation that x does or could take the place of y . This operative concept demands that one distinguishes between the respective positions that x and y hold, on the one hand, and their actual being filled in by x or y , on the other. Furthermore, it is implied that all positions that are held by some of the x s or y s are isotopic in the above-explained sense, and therefore, there can be in principle *various* assignments of places to the x s or y s in their common biographical situations. I do not, however, want to imply that only an *executed* change of position counts as a fulfillment of the pattern of representation. Rather, it suffices if each x or y that satisfies the behavioral standard of the pattern perceives itself as one case of filling in the occupation (with possible others) of such an isotopic position, which, at the same time, is related to many other isotopic positions.

This pattern of representation in all variations of biographical behavior is not already to be understood *expressis verbis* as acting for or representing others and their interest. Rather, it is generally and often only implicitly an ethically normatively *neutral* accompaniment of others at other places in relation to one's own, at which one behaves biographically as one does. It means to have an eye for the multiplicity of isotopic places in the field in which I myself behave. Accordingly, to being able to *represent oneself* in this sense means not merely to utter and pursue one's own interest immediately, but rather to pursue it *as* an interest of someone in my position or place. This place, however, is determined only in relation to many other isotopically occupied places in my environment. To *co-represent* these places in how I behave is insofar as almost inevitable. This means that I “*have*” to represent the other relevant ones if I “*can*” effectively represent myself and my interest in general. A small baby has and utters interests as well, but does so immediately, not in terms of a pattern of representation, as the interest of someone at this place. But already small children switch *de facto* into this mode, and in my experience much earlier than one could think at a first glance: If a toddler protests that she has only received one scoop of ice cream although her older sibling has received

two, and in similar situations, in which an interest is not only being uttered but also represented as the one of someone at a place which stands in relation to other places. For the pattern of representation as a standard of behavior fosters the biographical life of all participants, even if it can, in individual cases, equally be used for exploitation and especially egoistic goal pursuits, and is certainly often used in this way.

The propositions (P3) and (P4) taken together, that is the hold and widespread perception of the isotopy regarding all biographies and the biographical patterns of representation, which manifest themselves automatically at many places, amount to *the general form of life*, which is, according to my thesis, *ontologically*, that is *in* the thereby designated issue, responsible for the fact that all isotopically living individuals in such a live-association *are* persons without exception. They *are* persons in virtue of this form of their life that arises by itself, given appropriate conditions, as described in the four propositions. It is a *form* that does not come into being in virtue of the *nature* of the individuals but in virtue of their biographical – but under the respective burden – arbitrarily varying life. The form is finally *exercised* by those who are alive. But they cannot cease and refrain from doing so – except perhaps by externally induced interventions in their way of life. Furthermore, it is a form that cannot be instantiated by single individuals but only by certain life-associations of living individuals if the characteristic conditions of the four pivotal propositions are met. Therefore, persons only exist in the plural. This is – at least according to my intention – an *ontological* foundation of personhood that is *not* natural, and *not* “speciesistic”, but also *not* bound to the self-consciousness of each and every one who is a person. This means that persons are what they are indeed *by* nature but not *in virtue of* their nature, but due to what they always do in a formal sense, and as it were secretly – without being able to refrain from it.

Finally, I want to highlight some consequences of this ontological conception of personal existence for our own situation: It is clear that, following this, *all* human beings are persons, if humans live according to a general conviction in such a way that they meet P1 up to P4. That is, if, for example, an anencephalic child is born, then it is a person. It is, however, controversial which conditions need to be met to be able to speak of a complete *individual* case of human life. I do not have to say anything about that, since I have presupposed this for the entire conception. It is also clear that individuals that are *not* alive (like, for example, dead people) do not count as persons, except they would be considered alive. The question of whether there are actually non-human persons in our horizon according to the explained theory or not (e.g., certain primates) is in my opinion very difficult to assess, and according to what I have said consequently not to be excluded *a limine*. We would at least, where we had any cause of suspicion (even with artificially created beings), closely study the way of life of these beings if we did not want to go *wrong*. Is it a case of living beings at all? And is its way of life sufficiently biographically differentiated, and does it exhibit despite these differences certain

prevalent conformities? Are the standard cases of those beings disposed in such a way that they deploy the necessary *abstract-general* perceptions in regard to their isotopic life-situation, and that they also realize them in all their behavior biographically? And so on. It is, however, not the task of a philosophical conceptual analysis to exclude the possibility to misjudge something very easily. Even we humans have often enough *mistakenly* denied ourselves to be persons. For my proposed conception, however, it is specific that one can do something *wrong* at all regarding this question, and that it is not, as a last consequence, a question of our choice and ethical-moral convictions. The question whether certain occurring beings are persons or not is, according to my explained conception, an ontological one, and as such a question that needs to be answered according to significant criteria, even if such criteria are not easily empirically or with absolute certainty accessible to us.⁸

Notes

- 1 Spaemann (2006), pp. 9, 77, 232. Similarly Guardini (1952), p. 143.
- 2 Here I can only present the propositions for discussion in a sketchy form and explain their respective sense. I have argued for both positions at greater length at another place: See Buchheim (2018).
- 3 Thompson (2012), pp. 56–73.
- 4 Even Thompson (2012) concedes that our judgments (“natural historical judgments”) need, if they are true, a grounding in some matter of fact, since life forms in his sense are phenomena “bene fundata” (p. 76).
- 5 See, for example, Spaemann (2006), p. 245.
- 6 Spaemann (2006), for example p. 185: “They form a system of relations in which each is uniquely situated in relation to every other.”
- 7 See Spaemann (2006), p. 69.
- 8 I am grateful to Jörg Noller for translating the text into English.

Bibliography

- Buchheim, Thomas (2018): “What Are Persons? Reflections on a Relational Theory of Personhood.” In: Jörg Noller ed.: *Was sind und wie existieren Personen? Probleme und Perspektiven der gegenwärtigen Forschung*. Münster: Mentis: 31–55.
- Guardini, Romano (1952): *Welt und Person*. Würzburg: Werkbund.
- Spaemann, Robert (2006): *Persons: The Difference between ‘Someone’ and ‘Something’*. Oxford: Oxford University Press.
- Thompson, Michael (2012): *Life and Action. Elementary Structures of Practice and Practical Thought*. Cambridge, MA: Harvard University Press.

Part 2

Subjectivity

6 Personal identity and the transfer of commitment

Erasmus Mayr

1 Introduction

John Locke famously thought that the term “person” was a “forensic term”: Being a person was essentially a matter of being someone who could be held responsible and be called to account for what she did. Locke went on to provide an equally “forensic” argument for his view that connections of consciousness or memory were constitutive for diachronic personal identity. This argument turned on the question of which actions we could fairly be punished or rewarded for, especially by God at the day of the Last Judgement.

This personality extends it *self* beyond present Existence to what is past, only by consciousness, whereby it becomes concerned and accountable, owns and imputes to it *self* past actions ... therefore whatever past Actions it cannot reconcile or appropriate to that present *self* by consciousness, it can be no more concerned in, than if they had never been done: And to receive Pleasure or Pain; *i.e.* Reward or Punishment, on the account of any such Action, is all one, as to be made happy or miserable in its first being, without any demerit at all.

(Locke, Essay concerning Human Understanding, p. 346 f.)

Whatever you may think of Locke’s way to spell out what is needed for just punishment or reward, it seems hard to deny his basic idea that there is an essential connection between personal identity over time, on the one hand, and moral responsibility and accountability for earlier actions, on the other hand. At least, in the default case, personal identity transfers moral responsibility for earlier actions.¹ If James today is the same person as James★ in 2015, then *ceteris paribus* James can be held responsible for what James★ did back then (assuming that James★ was responsible for his actions at that time).

In this chapter, I want to examine another aspect of the general idea that transtemporal personal identity is tied to a transfer of accountability over time, *i.e.* that our present actions and stances give rise to normative commitments and requirements not only for us now, but also for us in the future. The connection between these two things goes far beyond the restricted “forensic”

idea Locke himself was interested in. It seems true for many attitudes and commitments that we are normatively accountable – be that epistemically accountable, rationally accountable, or morally accountable – both for and, as it were, “to” our own earlier acts, judgements, and propositional attitudes. When I say that we are accountable “to” our earlier acts and judgements, I mean that many commitments we have undertaken earlier continue to exert normative or rational pressure on us as long as we are the same person. Call this feature “normative transfer”. The notion of “normative” at issue here is a fairly minimal one, which lacks any moral connotations. Also “merely” rational commitments and requirements are “normative” in that attenuated sense, because they provide a standard to assess your behaviour and attitudes such that you can fail to meet this standard and there is something problematic if you do so.

This transfer holds only in the default case, for we may become aware of compelling reasons to change our minds; but absent any such changes, many of our earlier normative commitments “stay with us”. So continuing personal identity, in the default case, transfers responsibility *and* (and least a lot of) normative commitment – it is the paradigm way for such a transfer to happen. This is, I take it, a fairly trivial fact about personal identity. But it raises – or so I will argue – a considerable explanatory challenge for theories of personal identity of a reductive stripe: These theories should be capable of explaining why this normative transfer occurs.

In Section 2, I will say more about the kind of commitment which will interest me in the following, namely the one resulting from making decisions or forming intentions. In the subsequent sections, I will look at how two reductive approaches to personal identity could explain this transfer. My main focus (Section 3) will be on neo-Lockean accounts, and I will argue there are considerable grounds for scepticism about whether they can provide any such explanation. Afterwards, I will move on to an account of diachronic personal identity in terms of overlapping rational processes (Section 4), which, I think, has better prospects of explaining normative transfer, but faces significant problems, too, when construed as a reductive account of personal identity.

2 Commitment by intentions

It should be unsurprising that decisions and intentions – at least when they are about what to do in the future – give rise to commitments that “stay with us” over time. For on a natural understanding (at least future-directed) intentions either are or essentially include commitments to doing something later: When I now decide to do X tomorrow, and now form an intention with that content, I commit myself to doing X tomorrow. It is because of this feature that intentions and decisions can play their distinctive role in our motivational self-management which sets them apart from other motivational states such as desires. So, at least, goes one fairly plausible story about intentions and their function, which has been defended, e.g. by Michael Bratman.

In the following, I will be relying on some key elements of this broadly “Bratmanian” picture; while this picture is disputed,² for reasons of space I won’t be able to give any independent defence of the elements which are relevant for me here.

Future-directed intentions allow us to now settle the question of whether to do X later, i.e. to settle it before the time of action arrives. That we are able to make such in-advance commitments is essential for us as the finite and motivationally “fickle” kind of agents we are, as philosophers such as Bratman have stressed.³ Just to mention two crucial advantages this ability bestows on us: First, if I couldn’t make the commitment in advance, I would always have to decide whether to do X “on the spot”, i.e. when the action-situation has arrived. But then I may well lack the time, resources, and cool-headedness for adequate deliberation. So, it’s much better that I can do that now in advance, when I have the time to look for information and deliberate properly. Second, I may often fully believe (now) that doing X is what I should do tomorrow – but, at the same time, I may know that, if I left the decision until tomorrow, I would not then decide to do X, because tomorrow my desires will be very different from what they are now and I will feel strongly tempted to do something else or just feel too lazy to do anything. In such a situation, often⁴ the only way to ensure that I do X tomorrow (and thus to ensure that I will do tomorrow what I now think I should do then) is to now commit to doing X tomorrow – and that is precisely what forming a future-directed intention to do X allows me to do. Intention-formation thus allows us to deal with the fact that our desires are not generally stable, but all too often subject to change.

This latter feature, incidentally, shows that the “stabilizing” role of intentions cannot be restricted to intentions for future action. With regard to *any* activity – even ones we are already engaged in – which is extended over a significant period of time we face the same fundamental problem that we cannot be certain that our desires remain stable enough to continually motivate us to remain engaged in it and to perform each required part. If, at each step in such an extended activity, we were to deliberate anew whether to take this step and if following the overall activity would be completely “up for grabs” each time, we could not expect to follow through on most such activities.

The ability to commit now to a future course of action is therefore essential for us to deal with our limited resources for deliberation and our foreseeably changing desires. But intentions and decisions can only play the role of allowing us to settle practical questions in advance, if they have the following two features.⁵ First, they must have a certain measure of *de facto* stability, i.e. they must normally be fairly persistent and not something we generally give up “on a whim”. Second, they must have some *rational* stability: It cannot be rationally permissible to change our intentions *ad libitum* or to ignore them, just because we feel like it or just because circumstances have changed in some way or other.⁶ In particular, if the function of intentions lies, *inter alia*, in allowing us to pursue longer-term projects in the face of foreseeably

changing desires, then later changes in my desires cannot automatically justify abandoning my intention. This does not mean that I can never change my mind, whatever change in circumstances occurs; but I can only change it when the change in circumstances is “sufficiently important”, wherever that threshold lies.⁷ How precisely the rational stability of intentions works is a disputed issue;⁸ but the details do not matter here. What is important is only that there are certain constraints on rational intention-change, that these come from the characteristic function of intentions itself, and that they concern intention-change over time, i.e. they are diachronic requirements, rather than merely synchronic ones.⁹ That I have yesterday decided to go to the cinema this evening imposes rational constraints on me today: I cannot abandon the intention formed earlier without sufficiently good reason, and, if I haven’t abandoned it nor should abandon it, I am under a rational constraint to go to the cinema this evening. If I fail to do so (though I could), I have evinced a rational fault – I have, to some extent, been irrational.¹⁰

Both my reference to Bratman’s theory and the fact that I have formulated the above points in terms of the function and features of “intentions” may suggest that accepting these points relies on regarding (at least future-directed) intentions as mental states separate from the actions which issue from them. But it merits emphasis that this is not the case (nor do I want to commit to the latter view here). Talk of “intention” can be ontologically innocent: We can appropriately say that you “have a future-directed intention to do X” whenever you intend to do X later, where the latter formulation carries no ontological commitment to an independent mental state with it.¹¹ Even if you reject the idea that intentions are such independent states, the case for the claim that you should, *ceteris paribus*, stick to what you intend to do remains, in substance, the same. Take, e.g. Luca Ferrero’s view that there is no sharp or categorical difference between intending and acting; instead, both are “only different ways of describing distinct portions of a single, temporally extended process”, which begins, typically, with the adoption of a goal and goes on until the agent reaches this goal or abandons it.¹² On such an understanding of intending, the earlier considerations about our human limits and motivational “fickleness” still speak in favour of the view that you cannot rationally break off such a process without sufficiently good reason once you have adopted the goal.¹³

In committing myself for the future, I must think of myself as a temporally extended agent who is going to be the same agent, and the same person, in the future. Only thus can I commit *myself* to a future course of action. Put in terms of the intention’s satisfaction: My first-person singular intentions are necessarily about my own future behaviour and can only be satisfied by me. In understanding acting on an intention, we must therefore presuppose that the person who is acting is the same person as the one who formed the original intention. As Christine Korsgaard rightly put it in a critical discussion of Derek Parfit’s *Reasons and Persons*: “When the person is viewed as an agent, no clear content can be given to the idea of a merely present self”.¹⁴

3 Transfer of normative commitment and the neo-Lockean account of personal identity

3.1

Any viable account of diachronic personal identity (at least for human beings) must accommodate that insight and the fact that you can commit yourself for the future and later execute your own earlier intentions. Otherwise, a crucial element of our life as purposeful and deliberating, but finite agents would be lost (i.e. the role of intentions for future actions or temporally extended activities) – which would disqualify the account as a convincing account of our continuing personhood. But what is required for an account of personal identity to “accommodate” this insight will depend on what kind of account it is meant to be.

Theories of personal identity that lack reductive aspirations need not necessarily provide an independent explanation of why normative transfer for intentions occurs; they might, in principle, just take it on board as a (further) element of what it is or means to be the same person. For example, they may just claim that it is part and parcel of being the kinds of beings we are, i.e. finitely rational and motivationally “fickle” agents, that we are capable of forming intentions which possess the rational stability described.

This move is not available, though, to reductionist accounts of personal identity over time which try to fully explain continuing identity in terms of connections between different temporal stages of the person, where these connections are themselves not taken to be of a normative character but ones that can be described in purely descriptive terms.¹⁵ Assume that such a theory identifies features X, Y, Z as the features that must characterize the relation between Erasmus-Mayr-now and Erasmus-Mayr-yesterday for both to be temporal stages of the same person. Then we would need an explanation of why the presence of features X, Y, Z ensures that, at least normally, the normative commitments that were imposed on Erasmus-Mayr-yesterday by his intentions should be transferred to Erasmus-Mayr-today.¹⁶ That they are transferred is, I have claimed, one of the trivial and humdrum facts about personal identity we are all familiar with. But as long as no explanation was available, we would have to expect that the presence of features X, Y, and Z, on the one hand, and normative transfer, on the other hand, can easily come apart. Since personal identity and normative transfer cannot easily come apart, this would suggest that personal identity does not rest on the specified features X, Y, and Z after all.¹⁷

If we don't want to stack the deck unfairly against reductive accounts of personal identity from the start, we cannot demand that such an explanation of normative transfer includes that literally and nonderivatively the person who is acting later must be the same subject as the one who formed the intention in the first place. But we can demand an explanation for why normative commitment should be transferred between different temporal stages of

the person once the connections supposedly constitutive of personal identity over time obtain between those stages. To take a simple Lockean picture: If the relevant connections are to consist in memory-links, then there should be some explanation available of why the normative commitment should be transferred from the earlier to the later temporal stages when these memory-links obtain.

The transfer of normative commitment with regard to intentions has at least three features which will need to be accounted for.

- i First, my intention formed at t_1 puts me under the aforementioned rational pressures at t_2 , unless I change the intention in a rationally permissible way.
- ii Second, I am later capable of doing what there is rational pressure for me to do – discharging my rational obligation, as it were – by acting on that intention. And note: Doing X when I intended to X is not sufficient for acting on that intention. The latter requires doing it because I intend to do it/decided to do so. Imagine that I now intend to take a walk in the English Garden in the afternoon. But, unfortunately, I forget all about my intention during the discussion of my paper. However, I later absent-mindedly stroll through Munich, and inadvertently end up in the English Garden after all. In that case, I performed the kind of action I intended to perform (I took a walk in the English Garden). But I didn't satisfy my earlier intention. Doing so would have required that I had taken the walk because of my earlier intention or in the course of executing it.
- iii Third, if normative commitment is really transferred, then not only what there is normative pressure for me to do (i.e. the content of my commitment) remains the same, but also the grounds for this normative pressure remain the same, or at least, there is sufficient overlap, such that the grounds for the earlier commitment play an essential and non-redundant role in explaining the later commitment. To illustrate this point with an example where the normative pressure arises from interpersonal commitment: If I am now the same person as Erasmus Mayr★ one week ago, and if Erasmus Mayr★ was under normative pressure to go to Munich because he had promised to attend a conference, then Erasmus Mayr now (i.e. me now) should also be under that normative pressure because Erasmus Mayr★ (i.e. me back then) had made the promise earlier. It would not be enough if the normative pressure arose from some independent factor, e.g. from a new promise to go to Munich I made yesterday; in the latter case, the normative commitment wouldn't have been transferred, but would have been created anew later.

3.2

Let us now turn to the most prominent attempt to conceive of personal identity over time in terms of psychological connections between different temporal stages, i.e. the neo-Lockean approach, and examine whether it can

provide an explanation of the kind we are looking for. Broadly speaking, according to such a view, my present self belongs to the same person as an earlier self, if there are sufficiently many relevant psychological links between the two¹⁸ – e.g. retained states of consciousness and propositional attitudes, as well as connections in the form of cross-temporal references and causal links between these states.¹⁹ Locke himself thought that the relevant psychological connections were memory connections; but the general framework can be fairly liberal about which psychological connections are relevant for personal identity.²⁰ In particular, *practical* mental states, like intentions or general policies, and the presence of the right interrelations between them can be counted as equally essential for continuing personal identity. On such a natural extension of the original Lockean picture, X-now will only be the same person that Y-back-then was if X-now shares sufficiently many of Y-back-then's intentions, acts on sufficiently many of them and there are sufficiently many connections (both causal and referential) between X-now's intentions and actions and Y-back-then's earlier intentions.

This extension of the Lockean picture might, from the start, be suspected of introducing circularity. For my intention to do X is an intention that concerns *my own* future conduct and can only be satisfied by *my* doing X. So, whether a later self's doing X can count as an execution of the earlier intention, it seems, depends on the later self's being, or belonging to, the same person as the earlier one. The same would apply to retaining an earlier self's intentions. This worry is, of course, reminiscent of the well-known circularity worry about the memory criterion for personal identity. And just as in that latter case, neo-Lockeans are unlikely to be discouraged by it. To solve the worry for the memory case, a notion of quasi-memory, which does not already presuppose that the person who did X and the person who quasi-remembers doing X are identical, has been appealed to (e.g. by Sydney Shoemaker 1970/2008). The same move seems open with regard to intentions: While only I can satisfy my earlier intention to do X by doing X, there may well be a notion of quasi-intention such that another person's action could equally count as a satisfaction of my earlier quasi-intention.²¹ Let us assume, for the sake of the argument, that something along these lines works.²²

One sophisticated further development of the neo-Lockean view can be found in Bratman's work. While Bratman thinks that the described continuities and interrelations between states such as intentions are required for personal identity over time, he thinks we must further enrich this picture by bringing in additional higher-order attitudes. For, as Bratman notes, the continuities between our intentions are, at least in part, the "result of the agent's activity"²³: We can, at least to a certain extent, counteract motivational tendencies which would lead us away from our earlier intentions. This requires "monitoring and regulating my motivational structures",²⁴ which is the function of long-term plans and policies. The latter possess a "characteristic stability" which "normally induces relevant psychological continuities of intention and the like".²⁵ Furthermore, they are higher-order attitudes, i.e. attitudes towards other attitudes we have – such as a pro-attitude to ignoring

desires to engage in distracting activities when I'm engaged in writing a philosophy paper.²⁶ When these higher-order attitudes concern not just particular activities, or the management of desires, on specific occasions, but are commitments to recurring patterns of actions and/or the general motivational regulation of recurring desires and motives – such as a general commitment not to give in to one's desire for distraction when working – they are “self-governing policies”.²⁷ We will return to such policies below.

3.3

Does an enrichment of the original neo-Lockean model by connections between practical states allow this model to explain normative transfer with regard to intentions? What explains that, when two selves are parts of the same person, the later self is under rational pressure not to abandon the intention the earlier self has formed and is still bound by the commitment the earlier self has undertaken by forming her intention?

The simplest explanation on the neo-Lockean model might be the following: According to the extended neo-Lockean account, the earlier and later selves' intentions must overlap to a sufficient degree for both selves to belong to the same person. The later self must intend to do sufficiently many things the earlier self intended to do as well. But once the later self intends to do X, there seems an easy answer to the question why she is under rational pressure to do X: For, in virtue of now having the intention, she is rationally committed to doing X and, *ceteris paribus*, should do X.

But explaining normative transfer in this way is unsatisfactory, for the following reasons. First, it does not explain why normative transfer holds, *ceteris paribus*, for *all* intentions, whether or not they are actually retained or not. When I now form the intention to go to the cinema tonight, then I am under rational pressure to retain that intention, even if I end up abandoning it without good reasons. The suggested explanation, however, could not account for this rational pressure, since it can only account for a rational pressure once the later self, too, has an intention of the right kind. Second, as a consequence of this, the proposed explanation is not really an explanation of a *rational* pressure to retain the earlier self's intention at all. Rational pressure is something I can, in principle, fail to comply with. On the suggested account, this is not possible: Either my later self still has the intention – and has then complied with the rational pressure not to abandon the earlier one; or my later self lacks such an intention – and then it is under no rational pressure to have one because the commitment from the earlier intention no longer applies to it. Even if there is some causal mechanism in place which ensures that normally intentions of the earlier self are matched by corresponding intentions of the later one, this does not imply that there is something rationally criticizable when there is no corresponding intention of the later self. Third, the proposed explanation leaves out feature iii from the list of features of normative transfer from Section 3.1, i.e. the feature that the grounds for rational pressure should remain the same or at least significantly

overlap. According to the proposed explanation the rational pressure for the earlier self arose from *its* intention to do X; the rational pressure for the later self arises from *its own* intention. These are clearly different grounds. Even if there is some appropriate causal mechanism which connects both intentions, it is hard to see how its existence would make the rational pressure the later self is under depend in any crucial way on the earlier self's intention. For, on the proposed account, the rational pressure arises solely from the later self's having the intention now, and *this* pressure will obtain no matter whether the intention causally derives from the earlier self's intention or not. So, an appeal to the earlier self's intention will be redundant in explaining the rational pressure in question.

To meet this latter objection, neo-Lockeans who are sympathetic to the idea that we can explain the rational pressure in terms of the later self's intentions have to go beyond a merely causal connection between the intentions of the earlier and later selves and include some justificatory relation as well. One way to give the earlier self's intention such a role would be to posit that, in the case of personal identity, the later self must still *feel* bound by the earlier self's decision or intention. But this strategy would not explain full normative transfer, either. To begin with, feeling bound is not the same as actually being bound. It is not, to use a term of Korsgaard's, the normative "glue that holds an agent together"²⁸; it is rather the mere impression of normative glue. So, we would need some additional step from the feeling of being bound to really being normatively bound. Such a step, the neo-Lockean might argue, is provided by some principle such as "if you feel bound by a normative commitment, *ceteris paribus* you are bound". That is, a principle licensing the transition from feelings or seemings to the judgement that you are normatively committed. But even if such a principle were correct, it would not explain feature iii of normative transfer. The grounds for the earlier self's commitment were provided by its intention. The grounds of the later self's being committed, by contrast, are that it feels committed, and from that impression it can conclude that it is. These grounds are different, nor do they overlap in the required way: For where the "feeling of commitment" comes from is irrelevant for explaining the later self's commitment.

3.4

These considerations speak strongly against the possibility of explaining the phenomenon of normative transfer for intentions merely by appeal to corresponding intentions of the later self. But this leaves it open that other features of the later self might satisfactorily explain the rational pressure she is under. How could this work? Let's look at three possible strategies.

3.4.1

First, the neo-Lockean could argue that, in asking whether the later self is under rational pressure to do X, we should not only look whether she has an

intention to do X, but also take into account her other intentions and plans. The later self will be under rational pressure for coordination, integration, and coherence between her different intentions and plans.²⁹ Since she will share most of the other intentions and plans with the earlier self, if they are both (stages of) the same person, she is therefore under rational pressure to make her answer to the question of whether to do X coherent with (most of) the other intentions and plans the earlier self had. And the best way to at least avoid further inconsistencies, it seems, is to stick with the earlier self's intention to do X. For, assuming the earlier self had succeeded in coordinating her intentions and plans, dropping the intention to do X threatens to create new tensions.

It is hard to deny that there is some rational pressure of the kind this argument suggests. But it is insufficient to explain normative transfer. To begin with, the connection between the grounds of the rational pressure on the later self not to abandon the intention and the grounds for the rational pressure on the earlier self would still be fairly indirect. But even if this worry is set aside, the pressure this argument explains is much weaker than a rational pressure to stick with one's intention to do X (absent sufficiently good reasons to the contrary). For our different intentions and plans are not, realistically, all so tightly meshed that abandoning one of these intentions would generally lead to tensions with other plans or intentions. Such tensions may arise especially if large-scale projects are concerned where the planning necessarily involves significant adjustments of and "interweaving with" other plans. But many cases are not like that: For example, when I have planned to go to the cinema tonight, and decide at the last moment and without any reason to stay at home and do nothing instead, no such tension need arise.

3.4.2

A different neo-Lockean response would refer to the richer psychological structure that Bratman has described (*vide supra*). Since self-governing policies, as Bratman understands them, are meant to be one important factor that underlies the *de facto* stability of intentions, they may be promising candidates for explaining the rational stability of intentions, too. This seems particularly true for those self-governing policies which are not merely policies (not) to be motivated by a certain kind of incentive, but "to treat a desired end E as a justifying consideration",³⁰ i.e. policies about what to regard as justifying reasons in our practical deliberation.³¹ Such policies seem the right type of attitudes to explain not just why we usually don't change our intentions, but also to explain why we are under some rational pressure not to do so. Of course, appeal to self-governing policies does not explain the normative transfer for intentions all on its own, since for these policies, too, the question can be raised why we should stick with them. But a neo-Lockean-cum-Bratmanian could posit that an especially important psychological link for personal identity over time is the retaining of such policies which prohibit intention-change *ad libitum* with regard to those intentions the earlier self has held. Retaining such

policies would rank very highly among the psychological connections which determine whether the earlier and later selves belonged to the same person. Once the later self has such a policy and has no reason to give it up, there seems some rational pressure on it not to abandon intentions of the earlier self *ad libitum*. Otherwise, it would act contrary to its own self-governing policy and exhibit a form of practical incoherence.

However, this suggested explanation faces problems of its own. First, the self-governing policies Bratman refers to presuppose such a high amount of self-reflection and psychological complexity on the agent's part that it is fairly implausible to suppose that they govern our ordinary activities and everyday deliberations.³² At best, we are governed by such policies on rare occasions. Second, even when the later self has such policies, the fact that they support retaining an intention in light of temptations does not mean *per se* that the later self is under rational pressure to retain the intention. It might be equally rationally permissible for her to resolve the conflict between the policy and the desire to give in to the temptation by dropping or giving lower priority to the policy. Bratman himself raises this issue when he tries to explain the special authority possessed by self-governing policies which allows them to "speak for the agent" and present the agent's perspective. As Bratman notes, in cases of conflict between the policies and conflicting desires, "[w]ithin the agent's perspective, the self-governing policy seems to have a *presumptive normative authority*".³³ This authority, he believes, can be explained by the crucial role these policies play for our temporally extended agency and our continuing identity as *the same agent*:

so long as we are talking about agents like us, the agent of the action is an agent who persists over time and whose agency is temporally extended. [...] this is a deep fact about the kind of agents we are; or so I have averred. So it is natural to understand the relevant, agential perspective as the perspective of the temporally persisting agent whose agency is temporally extended.³⁴

But this explanation strikes me as circular within a neo-Lockean framework where we start out with distinct temporal stages or time-slices of persons which are meant to be "tied together" into one continuing person by psychological connections. Assume it is true that self-governing policies make temporally extended agency possible and thus have a better claim to constituting the temporally extended agent's "perspective" than the desires she experiences at this moment. Why should the agent-at-t, understood as the temporal stage of the agent at this time, give precedence to this overarching perspective over her merely momentary one?³⁵ If the agent-at-t already knew that she was part of a temporally extended person of which the earlier self in question was also a part and which would have future selves "belonging" to it, we might understand that she should care specifically about the overarching standpoint of that (persistent) person. Both because the earlier self's

projects were in some sense already “her own”, and because by adopting the overarching standpoint she could extend her own agency into the future. But whether the agent-at-t will be part of a temporally extended person, and whether she will be part of the same person as that earlier self, depends, on the neo-Lockean-cum-Bratmanian picture, in part on whether she and her future selves stick to sufficiently many earlier policies and give them precedence over their momentary desires. If they don’t, the present self will be part of a different person than the earlier self or may not itself have future selves. Thus, in advance of the present self’s settling whether to prioritize self-governing policies over momentary desires or not, it is open whether she will be part of a person whose temporally extended agency connects with the earlier and subsequent selves. But then appeal to temporally extended agency seems ill-suited to explain why the present self should give precedence to those policies.

It seems to me that once the possibility of a purely momentary perspective the agent-at-t could take is admitted at all, this problem arises *perforce*. The way to avoid it is to already reject the possibility of a momentary perspective as opposed to the overarching one. Persons, or at least persons like us, not only have the agential, temporally extended perspective available to them; this is the *only* one they have. *Prima facie* momentary perspectives are either overarching perspectives in disguise, or only possible within such a perspective. In support of the latter claim, I just want to note that even when I decide to act only on the desires I have now, this is not a decision I am taking within a purely momentary perspective: For I thereby necessarily take a stand on my earlier commitments. Even if this stand is one of rejection, I am necessarily relating to them in some way or other, which could not happen in a perspective which was truly momentary. But if it is correct that a momentary perspective is not truly available to beings like us, this seems to doom the neo-Lockean project which starts out with a series of selves which are, at least, initially independent from one another. That is one plausible way to understand Korsgaard’s assessment which I quoted earlier: “When the person is viewed as an agent, no clear content can be given to the idea of a merely present self”.³⁶

3.4.3

A third strategy would be to enrich the neo-Lockean account by introducing a more demanding *kind* of psychological link for personal identity over time. One option is to bring in the resources of narrative accounts of personal identity here. Jeanine Weekes Schroer and Robert Schroer, for instance, have argued that personal identity over time requires “narrative connectedness” such that the later self must have “a narrative explanation of at least some of the mental states/actions involved in the relevant relations/connections of psychological connectedness”.³⁷ Their resulting account of diachronic personal identity is still a reductivist one (and avowedly so), and also a reductivist one

of the kind that we are discussing here, since they do not take the narrative connection to be or to necessarily involve a normative connection.³⁸ Would such a further development of the neo-Lockean view lead to an explanation of normative transfer?

Two features of the account Schroer and Schroer offer make this very unlikely. First, they do not think that personal identity over time requires narrative connectedness to all mental states of the earlier self, but only to sufficiently many.³⁹ On their view, there will thus be intentions of the earlier self to which the later self is not narratively connected. As a result, their account will not give us any extra resources to explain normative transfer for these intentions. Second, even for those intentions that are narratively connected, it is hard to see that this fact would lead to any rational pressure on the later self to stick to those intentions. For having a narrative explanation for an intention does not imply that the subject sees any reason to endorse or maintain that intention; the intention could also be “narratively understood” as, e.g., an expression of a character-trait you are struggling with and are trying to reject.⁴⁰

Of course, the list of neo-Lockean responses I have reviewed here is far from exhaustive, and nothing I have said excludes that some other neo-Lockean explanation of normative transfer for intentions can be found. But in light of the problems I have sketched so far, I believe the prospects are not too good.

4 Normative transfer and an “overlapping processes” view of personal identity

While the neo-Lockean approach is normally understood as starting out with mutually independent temporal stages or time-slices of the agent, asking how these must be connected to form one continuous person, a different approach would start out with processes and activities,⁴¹ which are temporally extended and mutually overlap, such that different processes are ongoing at the same time. Since personal identity is at issue, the most relevant processes, it seems to me, have to be rational and (self-)conscious processes and activities of the right sort (such as activities of thought, deliberation, acting).⁴² A-at-time- t_1 would then count as (belonging to) the same person as B-at-time- t_2 , if there was a continuous series of sufficiently many temporally overlapping rational and conscious processes connecting both.

Such an approach is promising with regard to explaining why personal identity and normative transfer are connected for the following reason. We seem to have a grasp on something very much like normative transfer and on normative constraints through time in cases which do not, at least on their surface, rest on any presupposition of personal identity. One paradigm case is a joint rational activity whose different steps are distributed among different persons. Imagine that a group of people is reciting a poem in the way that each of them recites one line and then they move on to the next person. Within that one joint performance of the poem, the earlier steps normatively

constrain the later ones: When one person has recited line 6, the next person should recite line 7, and something goes wrong when she recites line 8. So the earlier steps in this performance constrain how one can correctly go on later. They provide a normative benchmark for these later steps. And they do so, even though we wouldn't suppose that this group of joint performers was one person through time. The same applies in cases of joint deliberation which is organized as a "round-game" argument, where each person is adding a further step in the argument. Within such a "round-game" argument, the earlier steps in the argument seem to create normative constraints on the later steps in just the same way as when it is one single person who is deliberating.

We can even drop the direct involvement of persons, when, e.g., we take a computer program going through one calculation. Nor does the constraining of later steps by earlier steps seem to require any underlying physical continuity; we can imagine that the program is transferred to another computer in the middle of the calculation, or the program copied and the calculation picked up at one point by the copy.

Thus, we seem to have a notion of normative constraints over time (later steps can go wrong in the activity or process, and go wrong because they don't rightly fit with the earlier steps), where this notion does not presuppose continuing personal identity of those engaged in this process. Even this notion, though, as illustrated by the examples, depends on *some* other notion of identity, namely of the identity of particular processes and activities. For the diachronic constraints I have mentioned earlier only apply within one and the same process. Think again of the joint recital of the poem. Obviously, the earlier steps (which lines have already been recited) only constrain the later ones if we presuppose that the people doing the recital are still engaged in the same performance and have not started a new one. If they have started a new recital, even of the same poem, the earlier steps don't impose any such constraints. That one person has already recited line 1 in the earlier recital is a reason not to recite this line again within the same recital, but no reason not to recite it again in another recital. Thus, our understanding of normative constraints within an activity presupposes an understanding of the identity of particular activities or processes. It is disputed whether we have any such understanding, especially when the processes are not yet complete, but still ongoing. Some philosophers, though, have argued that we can make sense of the identity of particular processes (e.g. Steward 2013), and, for the sake of the argument, I will assume here that they are right.

If they are, one could try to explain normative transfer as connected to personal identity in terms of diachronic normative constraints within *one* activity or process roughly as follows. If personal identity consists in or depends on a series of overlapping rational activities or processes and if there are diachronic normative constraints within the latter, then normative transfer on the personal level might be explicable as either a combination of these more local diachronic constraints or a generalization from them.

I believe that focussing on the structure of continuous rational processes and activities and the normative constraints arising from this structure is indeed helpful in understanding many aspects of the general phenomenon I have called “normative transfer”. But I am skeptical about whether we can, in this way, get a full explanation of this phenomenon, which could be part of a reductive theory of personal identity. I will briefly state my main worry.

It is far from clear how identifying particular processes works, and in particular, whether we can determine the identity of one process without some (at least indirect) reliance on the identity of underlying objects. One natural understanding of activities and processes is in terms of the exercise of underlying powers or dispositions of the objects or stuff involved in the process. On such a view, it is fairly straightforward to assume that the identity of a particular activity or process depends, *inter alia*, on the identity of the objects whose powers or dispositions are being exercised.⁴³ When, say, I am singing the first strophe of the Marseillaise and you are, directly afterwards, singing the second, then unless there is some coordination or mutual influence between our activities, it is natural to think that these are two numerically distinct activities, simply because the subjects who are exercising their capacities to sing the Marseillaise are numerically distinct for each strophe. If that is correct, it spells trouble for a reductive “overlapping processes” view of personal identity. For the relevant activities for diachronic personal identity will be rational and (self-)conscious ones, and the suspicion is that the only things that possess the capacities for engaging in these relevant activities are – well – persons. This would make the identity of these processes over time depend, *inter alia*, on the diachronic identity of the persons involved – which would make it circular to explain personal identity, in turn, in a way which presupposed the individuation of particular rational processes. This is of course no proof that we must end up in such a circle. But there is a genuine worry that we will.

The three cases I have presented earlier as cases of diachronic normative commitment without continuing personal identity don’t really answer this worry. The first two cases are cases of interpersonal cooperation. Even though different persons can perform the single steps of the overall activity, they must be members of the cooperating group. And the group’s identity over time depends, it seems, either on the identity of its members over time or (if these can change) at least on a sufficient overlap in their identity during sub-stretches of the overall process. The third case is one of artefacts, where you might worry that this brings in indirect reliance on the producer’s intentions when we talk of “constraints” on how the computer can proceed.

This is clearly not a knockdown argument against the possibility of an “overlapping processes” account, nor is it meant to be. But I hope it indicates why explaining normative transfer for intentions is a major challenge for such an account, and why it is far from clear how this challenge can be met.

5 Conclusion

Personal identity over time is connected to what I have called normative transfer for intentions: The commitments we form by making decisions and forming intentions for the future “stay with us” over time, in the default case. Reductive theories of personal identity should be able to explain this connection. But as I have argued for the neo-Lockean approach and an “overlapping processes” account, we should be fairly skeptical that they will succeed in doing so.⁴⁴

Notes

- 1 Maybe there are exceptions to this rule: Perhaps something like radical moral conversion rules out that you can still be held morally accountable for things you did earlier, even though you are the same person. But this case is not only debatable; even if responsibility and personal identity come apart here, it is certainly a special case, which does not put into question the fact that, in the default case, both go together.
- 2 See, e.g., Tenenbaum (2016).
- 3 See Bratman’s “planning theory of intention” in his (1987).
- 4 It need not be the only way: Sometimes, devices like “Ulysses contracts” are available alternatives. But too often such devices are not available or unfeasible.
- 5 For these features, see Bratman (1987), 16 ff., and his (2007d), 263 f.
- 6 This rational stability may be restricted to intentions where there is “both initial, supposed support for that intention and constancy of view of the grounds for that intention” as Bratman (2012), 78, suggests.
- 7 In Mayr (2017), I argue that these considerations support the “guise of the good” thesis about acting on intentions.
- 8 In particular, there is the question of whether we are rationally required merely not to abandon or even not to reconsider an intention without sufficiently weighty reasons (see Bratman 1987, chs. 5 and 6; Holton 2009, 121 ff.).
- 9 Of course, the existence of genuinely diachronic rational requirements has been disputed, see, e.g., Hedden (2015); but, for reasons of space, I cannot offer any independent defence of it here.
- 10 One may wonder whether this is true in every case, or whether we should just stick to our intentions “as a rule” (though we can, without irrationality, drop them on occasion even without any good reason). I will assume the former, stronger principle here. For a defence of this view see Bratman (2012).
- 11 As Ferrero (2017), 32 f., rightly points out.
- 12 Ferrero (2017), 14.
- 13 In fact, Ferrero himself makes a similarly ‘irenic’ suggestion about his view and key elements of Bratman’s planning theory of intentions (2017), 31.
- 14 Korsgaard (1996), 372.
- 15 I will not be discussing narrativism separately here, because it is not *per se* a reductivist approach. There are reductivist versions of this approach, though, e.g. Schroer and Schroer (2014). I will consider their view below as a further development of the neo-Lockean approach, since they locate their view explicitly in the framework of the psychological continuity approach, see (2014), 259.
- 16 Perhaps, strictly speaking, this explanation need not be provided by the account of personal identity all on its own, but only when combined with a plausible view of intentions. I will understand future-directed intentions as commitments which I (or my present self) now undertake(s) to do something later (as the same

- person). This characterization of future-directed intentions seems harmless enough and sufficiently plausible that, at least *prima facie*, a failure to explain normative transfer will be the fault of the theory of personal identity in question rather than of this characterization of intentions.
- 17 This argument is similar to the one offered by Schechtman (2007), 164, who first identifies four key features of personhood and then demands that “a satisfying psychological account should thus define identity in such a way that the relation that constitutes identity supports the four features”. I just think that normative transfer for intention is yet a further key feature that should be explained.
- 18 As many adherents of the Lockean picture have pointed out, it is not really necessary that two selves be thus connected for both to belong to the same person; it is sufficient that there is a continuous series of selves linking one to the other, such that between any subsequent selves there are sufficiently many connections of the right kind (see Noonan 2003, 10). But this modification will play no role for our following arguments.
- 19 Presumably, this picture, in the late 20th century, has been most prominently connected with Derek Parfit (though he only claimed that psychological continuity was “what mattered in survival”, not that it was sufficient for continuing personal identity (1971/1975)).
- 20 See Noonan (2003), 10.
- 21 Parfit (1971/1975), 210, assumes, without much further argument, that there must be such a notion of “q-intention”, parallel to “q-memory”.
- 22 An alternative answer to the circularity worry would be to construe the relevant first person “I”-intentions as “we”-intentions, (on “we”-intentions in general see Tuomela 2007). This construal is suggested by some remarks of Christine Korsgaard’s about the necessity of cooperation between present and future selves. See her (1996), 372, or (2009), 202 ff. (This is not to say that Korsgaard herself subscribes to the neo-Lockean project; indeed, in the first passage just referred to, she goes on to argue that already the assumptions of the cooperation picture are flawed.)
- 23 Bratman (2007b), 30.
- 24 Bratman (2007b), 30 f.
- 25 Bratman (2007b), 32.
- 26 See Bratman (2007b), 33.
- 27 Loc. cit.
- 28 Korsgaard (2009), 204.
- 29 See for this, again, Bratman (1987), and Mayr (2017), sec. 5.
- 30 Bratman (2007c), 103.
- 31 Bratman (2007c), 101.
- 32 See Hornsby (2004), 14 f.
- 33 Bratman (2007c), 104.
- 34 Bratman (2007c), 105.
- 35 Bratman suggest as an answer that the self-governing policies “play a central role in constituting the very perspective that is at issue in such talk about presumptive normative authority” (2007c), 105. I think this is correct, but it doesn’t yet show that the present self should give precedence to the policies, as long as she also has the momentary perspective (of ‘just-me-at-this-time’) available to her. If it does, then there will be two perspectives she can take: The perspective of the temporally extended agent and the momentary one, and it is difficult to find a decisive reason why she should prefer the first over the second. This problem can be avoided, I think, only by already rejecting the possibility of a genuinely momentary perspective which could be contrasted with the perspective of the temporally extended agent (see below), and I don’t understand Bratman as doing that.

- 36 Korsgaard (1996), 372. Though her argument is different because it is based on the relation to your own future, I think she would agree with this reading of her conclusion.
- 37 Schroer and Schroer (2014), 459.
- 38 That is, at least, how I read them.
- 39 Schroer and Schroer (2014), 461.
- 40 Compare Schroer/Schroer's coffee taste example, (2014), 456.
- 41 Process views of personal identity can look very differently; see, for a comparison of two examples, Siakel (2014).
- 42 Adherents of what is called 'processual animalism' will, to a considerable degree, disagree. (See, e.g. Meincke (2020) for a recent defence of this view). But, for reasons of space, I cannot discuss this position here.
- 43 Insofar as particular objects are involved in the process, that is.
- 44 For helpful comments and discussions on earlier drafts of this chapter I am indebted to Stefan Brandt, Nora Heinzlmann, Christian Kietzmann and Konstantin Weber, as well as to audiences in Munich and Erlangen.

Bibliography

- Bratman, Michael (1987): *Intentions, Plans, and Practical Reason*, Cambridge/MA: Harvard University Press.
- Bratman, Michael (2007a): *Structures of Agency*. Essays, New York: Oxford University Press.
- Bratman, Michael (2007b): "Reflection, Planning, and Temporally Extended Agency", repr. in Bratman (2007a): *Structures of Agency*, 21–46.
- Bratman, Michael (2007c): "Two Problems about Human Agency", repr. in Bratman (2007a): *Structures of Agency*, 89–105.
- Bratman, Michael (2007d): "Temptation Revisited", repr. in Bratman (2007a): *Structures of Agency*, 257–282.
- Bratman, Michael (2012): "Time, Rationality, and Self-Governance", *Philosophical Issues* 22, 73–88.
- Ferrero, Luca (2017): "Intending, Acting, and Doing", *Philosophical Explorations* 20 (2), 13–39.
- Hedden, Brian (2015): "Time-Slice Rationality", *Mind* 124 (494), 449–491.
- Holton, Richard (2009): *Willing, Wanting, Waiting*, Oxford: Oxford University Press.
- Hornsby, Jennifer (2004): "Agency and Actions", in John Hyman and Helen Steward (eds.), *Agency and Action*, Royal Institute of Philosophy Supplement 55, Cambridge: Cambridge University Press, 1–23.
- Korsgaard, Christine (1996): "Personal Identity and the Unity of Agency: A Kantian Reply to Parfit", repr. in Korsgaard, *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press 1996, 363–398.
- Korsgaard, Christine (2009): *Self-Constitution. Agency, Identity, and Integrity*, Oxford: Oxford University Press.
- Locke, John (1975): *An Essay concerning Human Understanding*, P. Nidditch (ed.), Oxford: Oxford University Press.
- Mayr, Erasmus (2017): "Acting Sub Specie Boni", *Philosophical Explorations* 20 (2), 55–72.
- Meincke, Anne Sophie (2020): "Processual animalism. Towards a scientifically informed theory of personal identity", in Anne Sophie Meincke and John Dupre (eds.), *Biological Identity. Perspectives from Metaphysics and the Philosophy of Biology*, London: Routledge, 251–278.

- Noonan, Harold (2003): *Personal Identity* (2nd ed.), London: Routledge.
- Parfit, Derek (1971/1975): "Personal Identity", repr. in John Perry (ed.), *Personal Identity* (1st ed.), Berkeley: University of California Press 1975, 199–226.
- Schechtman, Marya (2007): "Stories, Lives, and Basic Survival: A Refinement and Defense of the Narrative View", *Royal Institute of Philosophy Supplement* 60, CUP, 155–178.
- Schroer, Jeanine Weekes / Schroer, Robert (2014): "Getting the Story Right: A Reductionist Narrative Account of Personal Identity", *Philosophical Studies* 171 (3), 445–469.
- Shoemaker, Sidney (1970/2008): "Persons and Their Pasts", repr. in John Perry (ed.), *Personal Identity* (2nd ed.), Berkeley: University of California Press 2008, 249–282.
- Siakel, Daniel (2014): "The Dynamic Process of Being (a Person): Two Process-Ontological Theories of Personal Identity", *Process Studies* 43 (2), 4–28.
- Steward, Helen (2013): "Processes, Continuants, and Individuals", *Mind* 122 (487), 781–812.
- Tenenbaum, Sergio (2016): "Reconsidering Intentions", *Nous* 52 (2), 443–472.
- Tuomela, Raimo (2007): *The Philosophy of Sociality: The Shared Point of View*, Oxford: Oxford University Press.

7 Biological subjectivity

Processual animalism as a unified account of personal identity

Anne Sophie Meincke

1 Introduction

Opposing the widely accepted Psychological Approach to personal identity, the Biological Approach, also known as ‘animalism’, holds that the ontological conditions of our identity through time are purely biological. As Olson (1997, 4) puts it: “no sort of psychological continuity, with or without further physical qualifications, is either necessary or sufficient for us to persist through time”. This is true, according to Olson (2015), because we are animals and the nature of animals is such that animals have purely biological identity conditions.¹ Animalism thus understood² is increasingly popular (see, e.g., Blatti and Snowdon 2016); but it faces challenges. One of these³ is the worry that animalism, by reconceptualising personal identity as biological identity, fails to capture the practical and ethical importance we attach to our identity through time since this appears to be linked to psychology (Baker 1999; Shoemaker 2016).

Olson is aware of this challenge. He recognises the attractiveness of “the idea that any acceptable account of our identity must explain the moral significance of our identity, and that we have whatever moral significance we have by virtue of being people” (Olson 1999, 165). Yet he rejects the hypothetical conclusion “that the conditions of our identity must derive from our being people”, insisting that it is by no means clear whether “relations of practical importance always coincide with strict numerical identity” (ibid.). In fact, Olson (1997, 52–72) argues that the latter is not the case and that we ought to consider complementing the ontological theory of numerical identity provided by animalism with a separate ethical theory of personal identity. This, he emphasises, is as good as it gets, given that “no account of our identity has yet been proposed that guarantees [...] the coincidence of what is important in our identity with the actual conditions of our identity” (Olson 1999, 165).

In this chapter, I take issue with Olson’s pessimism about the possibility of a unified account of personal identity, and with the explanatory patchwork he offers us instead. We can do better than that. More precisely, *animalism* can do better. We don’t have to deny that we are animals in order to overcome the supposedly unsurmountable gap between ontology and ethics. On the

contrary, animalism possesses all the resources needed for this if put on the right ontological footing.

I substantiate this thesis in four steps. First, I show how the envisaged division of labour between animalist ontology and psychological ethics is far less harmonious than Olson would have us believe. In fact, it gives rise to serious conflicts in cases where biological continuity and psychological continuity supposedly come apart. Second, I argue that the root of the problem lies in the exclusion of subjectivity from the animalist concept of an organism and of biological life. This exclusion of subjectivity turns out to be linked to the view that an organism, ontologically, is a substance: a biological substance, whose identity is independent of the possession of psychological properties. Third, I review recent scientific work from biology, cognitive science and the philosophy of mind that provides evidence for cognition and consciousness in a large number of non-human organisms, perhaps even in all organisms. Subjectivity is not tied to personhood but appears to be biologically founded in the organisation of living systems. Finally, I explain how to make sense of these empirical findings in metaphysical terms. Subjectivity can be integrated into a biological account of our identity through time once it is acknowledged that organisms – including human organisms – are processes rather than substances. According to the novel *Processual Animalism* that I propose, organisms are biological subjects in that they exist as the process of establishing and maintaining a self/non-self-distinction through the interactive construction of boundaries. As the Cartesian-style dichotomy between biology and psychology dissolves, the lamented gap is closed: ontology and ethics unite in biological subjects.

2 Animalist ontology versus psychological ethics: complementarity or conflict?

Olson (1997) proposes a Biological Continuity Criterion of identity through time:

If x is an animal at t and y exists at t^* , $x = y$ if and only if the vital functions that y has at t^* are causally continuous in the appropriate way with those that x has at t .

(Olson 1997, 135)

Our identity through time, like that of any other kind of organism, consists in the continuous, undisrupted operating of those biological functions that sustain our life.

The Biological Continuity Criterion vindicates the commonly held beliefs that each of us once was a foetus (Olson 1997, 73ff.) and may be a patient in a persistent vegetative state (PVS) one day (Olson 1997, 7ff., 111ff.). Olson regards this as a strong case for animalism and against the Psychological

Approach, according to which psychological continuity is both necessary and sufficient for personal identity.⁴ Since, as he claims, neither the foetus nor the PVS patients have any mental states (or, at least, any mental states sophisticated enough for personhood), they cannot be psychologically continuous with a person and, hence, cannot be numerically identical with a person. The psychological theories are forced to conclude that each of us comes into existence not before some point after birth and could not survive the loss of their mental life.⁵

Proponents of the Psychological Approach typically appeal to the so-called Transplant Intuition: if my brain were transplanted into another (brainless) body, I – the person – would follow my brain and continue to exist with a new body. ‘The person goes where the brain goes.’ This is assumed to be the case because, thanks to the continuous functioning of my cerebrum, there is psychological continuity before and after the transplantation. But Olson (1997, 42ff.) argues that the Transplant Intuition is mistaken (at any rate as far as cerebrum transplantations are concerned). The Transplant Intuition “leads us astray” (Olson 1997, 47) as soon as we consider the possibility of a person’s two hemispheres being transplanted into two different bodies because then two persons would be psychologically continuous with the pre-operative person, e.g., with me. According to the Psychological Continuity Criterion, we would have to conclude that both post-operative would be numerically identical with me; but this is logically impossible. Animalism (supposedly)⁶ avoids this difficulty. Whether my cerebrum will be transplanted as a whole into a new body or be split into its two hemispheres, which then are transplanted into two different bodies, the answer given by the animalist to my anxious question ‘Who will I be after the transplantation?’ is the same: I won’t go where my cerebrum goes. I will rather stay behind. I will be the animal that has been deprived of its cerebrum but that, thanks to the continuous functioning of my brain stem, is still alive.⁷ This follows from the Biological Continuity Criterion.

However, Olson (1997, 52ff.) admits that I, confronted with the prospect of losing my cerebrum, nevertheless have reasons to be concerned about the well-being of the person who ends up having my cerebrum and, likewise, about the well-being of the two persons who might end up having each one of my cerebral hemispheres. Exactly because there is *psychological continuity* between me and them, it makes perfect sense to wish them a happy life without hardships and misfortunes. Olson accepts what he calls the Parfit-Shoemaker Thesis, according to which “prudential concern doesn’t always follow strict identity” (Olson 1997, 54). So if I, confronted with the prospect of having my cerebrum removed and transplanted elsewhere, were to ask the different question ‘Whom should I be concerned about after the transplantation?’, the answer would be: ‘I should be prudentially concerned about the person who possesses my cerebrum, or about the persons who each possess one of my hemispheres, even though this person, or these persons, wouldn’t be me.’

Olson (1997, 56) surmises that the Transplant Intuition reflects such prudential interests and concerns:

We believe that one survives if and only if one's mental features are preserved because that is what it takes for someone to be worthy of our prudential concern, and because it is natural to suppose that prudential concern always coincides with identity.

Olson's point is that we are right to be prudentially so concerned and that this concern, as a valid principle on which the (otherwise erroneous) Transplant Intuition rests, can be accommodated in the animalist picture. We simply have to accept that where prudential concern and numerical identity coincide (which is usually the case), this is only contingently so. The same, according to Olson (1997, 57ff. and 63ff.), holds true for moral responsibility and for the way we may treat people within a given social context, which both turn out to hinge upon psychological continuity rather than numerical identity.

The upshot is that none of the usual arguments for the Psychological Approach actually support that view insofar as it is meant to provide an ontological explanation of our identity through time. Instead, Olson claims, it is largely "practical considerations" that motivate the Psychological Approach, and these "may well be compatible with the Biological Approach" (Olson 1997, 42). Accordingly, Olson proposes a revised reading of the Psychological Approach's identity claims in terms of an explicitly practical so-called 'being the same person as' relation:

x is at time t the same person as y is at a later time t^* if and only if x ought to be prudentially concerned, at t , for y 's well-being at t^* ; and y is responsible, at t^* , for what x does at t ; and it is natural and right at t^* to treat y as if she were x .

(Olson 1997, 66)

The 'being the same person as' relation presumably requires psychological continuity as a satisfaction condition – "roughly speaking, x is now the same person as y is later on just in case y is then psychologically continuous with x as she is now" – though Olson is prepared to leave the ultimate decision about this to the ethicists "because, *being the same person*, as we might say, is not a metaphysical relation" (Olson 1997, 69). Being merely practical, the 'being the same person as' relation thus can peacefully co-exist with the relation of biological continuity that, according to animalism, constitutes numerical identity. Once it is recognised that the Psychological Approach concerns personal identity in a practical rather than ontological sense, the traditional rivalry between the Psychological and the Biological Approach can give way to harmonious complementarity between ontology and ethics. Or so Olson argues.

Harmonious complementarity, surely, is not the same as unity. And is it even that harmonious? To “divorce numerical identity from those relations of practical concern that are traditionally thought to go along with our identity” (Olson 1997, 70) means to allow for the possibility that these two things actually can come apart because psychological continuity and biological continuity can come apart. Olson explains:

I might be morally accountable for something that someone else did, and not accountable for things that I did, even if I am morally competent now and I was morally competent then. That is because, on the Biological Approach, I might be psychologically continuous with some past person other than myself; and I might not be psychologically continuous with myself as I was at some past time.

(Olson 1997, 70)

Imagine that I, the human animal going by the name ‘Anne Sophie’, have my cerebrum replaced with that of some other female human animal called ‘Anne Marie’. Olson, in the above quoted passage, describes this situation as one in which I, Anne Sophie, am now morally responsible for what Anne Marie did in the past, and in which I, Anne Sophie, am no longer accountable for what I, Anne Sophie, did before the operation. But this is not how the ethicists, according to Olson’s own reconstruction of their motivations, would describe the situation. The ethicists, Olson told us, would – or, at least, should – apply the ‘being the same person as’ relation; and this would result in the judgement that the human animal that used to go by the name ‘Anne Sophie’ is now the same person as Anne Marie. Accordingly, they would treat this animal as Anne Marie, which includes corresponding moral judgements. So from the point of view of the ethicists, there is no mismatch between personal identity and moral responsibility. Anne Marie, the person, has been there all along; it’s just that Anne Marie now, so to speak, has a different body. But this does not hinder holding her accountable for what she did before she changed her body.

Olson’s description of the situation is given from a perspective that ties numerical identity to biological continuity. But the way this description mixes this proclaimed ontological perspective with ascriptions of moral responsibility obscures the fact that, in our case at hand, the ontological and the ethical perspectives, as set up by Olson, yield directly contradictory identity statements: for the ontologist, the post-operative person is Anne Sophie, who just happens to have a new cerebrum (namely Anne Marie’s); for the ethicist, the post-operative person is Anne Marie, who just happens to have a new body (namely Anne Sophie’s) – and it’s her, Anne Marie, who is responsible for her past deeds, not Anne Sophie.

Contradictions are not conducive to harmony and even less to unity. Of course, Olson might respond by stressing that the ontological identity statement is the one that ultimately counts because ontology is more fundamental

than ethics. But this move would just lay open the lack of argument when it comes to explaining the ontological priority of biological continuity over psychological continuity. Where does the relation of biological continuity get its ontological power from, and why does the relation of psychological continuity lack such ontological power, instead being merely practically relevant? No reasons intrinsic to these relations are given; psychological continuity is denied carrying ontological weight, it seems, solely on the grounds that if it were to be ontologically potent, this would lead to problems to do with possible reduplication, while no effort is being made to show that biological continuity does not face such troubles.⁸

3 The exclusion of subjectivity in substance ontological animalism

The possibility of contradicting ontological and ethical identity statements reveals a fundamental difficulty for animalism. Consider Olson's above quoted statement: "I might be psychologically continuous with some past person other than myself; and I might not be psychologically continuous with myself as I was at some past time". Who is 'I'? As we have already seen, it is not who Olson's ethicists – and, for that matter, the proponents of the Psychological Approach – would expect it to be; it is not the *person* whose psychological continuity is preserved despite her going through what the ethicists and psychological theorists interpret as a switch from one to another hosting body. With a view to our example: 'I' does not refer to Anne Marie; instead, Olson uses the first-person pronoun to refer to Anne Sophie. But who is Anne Sophie? Ontologically speaking in Olson's sense, Anne Sophie is the *animal* that loses its cerebrum and gets a new one from Anne Marie. This animal Anne Sophie, however, can refer to itself with the first-person pronoun only thanks to Anne Marie's cerebrum and, i.e., thanks to having Anne Marie's psychology.

In view of this fact that Anne Sophie can refer to herself only thanks to Anne Marie's psychology, one may indeed wonder whether she would identify herself ontologically correctly as Anne Sophie rather than Anne Marie. Maybe Anne Sophie would, following the operation, be systematically mistaken about who she is, stubbornly insisting on being Anne Marie. Of course, Olson might be fine with this. He might think that it is sufficient if the animalist metaphysicians know the truth about Anne Sophie's identity. But, then, what is this truth founded upon? What has the animalist to offer to the ethicist or proponent of the Psychological Approach who insists that, after the operation, the person Anne Marie has the better claim to using the first-person pronoun 'I' than the animal Anne Sophie has, given that Anne Sophie, without Anne Marie's psychology, could not use this pronoun at all?

The first-person pronoun 'I' is perspectival. It expresses a subjective point of view onto the world. But the animalist, by excluding psychology from being ontologically constitutive of human animals and their identity through

time, has deprived herself of what appears to be the intuitive basis for ascriptions of subjectivity. Yet, the animalist keeps speaking in subjectivity-laden language: ‘I was a fetus’, ‘I may be a PVS patient one day’, ‘I might be morally accountable for someone else’s deeds’. Animalism is declared to be a theory of “*our* identity through time” (Olson 1997, 7; italics added); it endeavours to demonstrate that “no sort of psychological continuity, with or without further physical qualifications, is either necessary or sufficient for *us* to persist through time” (Olson 1997, 4; italics added). What do the personal pronouns refer to when no subjects are present to refer to?

Olson does not deny that human animals can be persons and thereby subjects – they just need to satisfy the criteria of personhood specified by Locke (Olson 1997, 94ff.). But these personal features are temporary and contingent, whereas the human animal persists as the same throughout the coming and going of psychological states. To use an analogy: just as the same software can run on different hardware and the same hardware can host different software, so chains of psychological events may travel from animal to animal and the same animal may end up hosting different psychologies. The question then becomes how to handle hardware on which – not yet or no longer – any software is installed.⁹

Hardware without any software running on it is, however, the rule in biology, as described by Olson. “Many animals – oysters, for example – manage to persist without having any psychological capacities at all” (Olson 1997, 110). We are therefore well-advised not to stipulate special persistence criteria for those few animals who happen to have psychological capacities at least occasionally (Olson *ibid.*) considers dogs, pigeons and fish). This would lead to heterogeneous persistence theories in the realm of life (*ibid.*). Instead, we should regard ‘animal’ or ‘living organism’ as a so-called *substance concept*: a “concept that tells us, in a special sense, what the object [that falls under it; A. S. M.] is” and that “determines persistence conditions that necessarily apply to all (and perhaps only) things of that kind” (Olson 1997, 28; see also Wiggins 1980, 2001). The persistence conditions determined by ‘animal’ or ‘living organism’ are purely biological, which means that “[...] our persistence conditions are the same as those of aardvarks and oysters and other animals” (Olson 1997, 30). Moreover, *qua* substance concept, the concept ‘animal’ or ‘living organism’ applies to animals or organisms throughout their lives, whereas we should treat ‘person’ as a mere *phase sort*, applying only to certain sections of a human animal’s life, namely in those sections in which human animals have psychological states (Olson 1997, 30f., and 110).

If having no psychological states is the default in nature, is the problem of how to justify first-person subjective reference to human animals that lack psychological states (or to human animals apart from any psychological states they may have) then just an exceptional and, hence, negligible, problem? I don’t think so. After all, Olson (1997, 26f.) insists that animalism, although excluding psychology, still provides an account of *personal* identity rather than of animal identity. Animalism, as a contribution to the debate in metaphysics

on personal identity, is concerned most importantly with those animals that (at least normally and at least for most parts of their lives) are persons: human animals. Consequently, it would seem that the supposedly exceptional status of human animals within the animal kingdom does exactly not absolve the animalist from saying more about these personal aspects, including explaining what it is that justifies extending references to a *subject* via first-person pronouns from personal to non-personal phases or aspects of human existence. On the contrary, we deserve to know, if personality, or psychology more generally, is actually thought to be superficial (ontologically speaking), in what sense human animals can be proper subjects at all.

Metaphysics and ethics, in Olson's picture of our identity through time, dissociate as a result of psychology and subjectivity being expelled from the former's province and being confined to the latter's. Mere animals deprived of subjective experience are poor objects of personal identification, prudential concern and moral considerations. Hence, Olson is prepared to accept Parfit's provocative thesis that numerical identity is not what matters, leaving what matters instead to the Psychological Approach reframed as – or rather degraded to – a merely prudential theory. At the same time, however, animalism's missing basis for personal identification is concealed by continuous ontological references to human animals qua merely (i.e., psychology-and-subjectivity-free) biological beings through the first-person pronouns 'I' and 'we', and even by talk of 'persons' where – according to Olson's own animalist theory – there are none, as in the PVS case (Olson 2007, 40).¹⁰ There are thus two competing senses of 'personal identity' in Olson's animalist theory: on the one hand, 'personal identity' as a practical relation where 'person' is a role assumed by organisms at certain times and under certain conditions that are prudentially, morally and socially defined; on the other hand, 'personal identity' as the ontological relation of 'our' identity through time.

This paradoxical ambiguity and the unresolved problem of seemingly illegitimate first-person subjective references that comes with it indicate that Olson's animalism, contrary to its official agenda, cannot do without subjectivity. However, before appreciating why animalism actually does not need to do without subjectivity, we ought to understand better why Olson thinks it can. To this end, we need to dig a little deeper into the ontological ground on which animalism grows.

Recall Olson's claim that 'animal' or 'living organism' is a 'substance concept'. Substance concepts are sortals that categorise substances. Animalism conceives of organisms as substances (Olson 2007, 27). Substances are (as I put it) *things*: discrete particulars whose existence is independent of the existence of other things and whose identity at a time and over time is independent of the occurrence of change. A substance is what it is thanks to its intrinsic properties, namely those that make up its essence: properties the substance cannot lose without ceasing to exist. Which properties these are is specified by the substance sortal under which a substance falls. In the case of

organisms, according to Olson's animalism, only certain biological properties are essential. This, as we have seen, does not prevent at least some organisms, such as human organisms, from instantiating also a different type of properties: psychological properties. However, psychological properties belong to the class of merely accidental properties, which the organism *qua* substance can change while maintaining its identity. Human animals, just like other animals, are biological substances with psychological characteristics as non-essential add-ons.

Note that this picture contains only one type of substance: biological substances, which presumably count as a sub-kind of physical substances. Because the concept 'person', on Olson's view, does not qualify as a substance sortal, we are not dealing with substance dualism here. Yet, there is another sort of dualism, namely between two types of properties: physical/biological properties on the one hand and psychological/biological properties on the other. Though they tend to co-occur, they are ontologically strictly different, which is reflected in the fact that they *need not* co-occur. Physical/biological properties can occur without any psychological/mental properties co-occurring, and formerly co-occurring chains of physical/mental property instantiations and of psychological/mental property instantiations can part ways (at least in certain respects).¹¹ "The Biological Approach", Olson (1997, 126) explicitly states, "is intended to be compatible with a 'dual-aspect' or 'property-dualist' theory of mind, according to which psychological properties are in some sense non-physical properties".

The substance or thing ontological framework facilitates the idea that one can have an organism without psychology and subjectivity. If psychology and subjectivity are just properties to be instantiated by a substance, then it is not hard to imagine that the substance's existence does not depend on the instantiation of these properties. Quite the contrary, as with all properties, it seems there need to be a substance first for any properties to be instantiated. With the exception of a substance's essential properties, the substance's existence and identity are logically and ontologically prior to any properties, which it might instantiate. Of course, we could also conceive of substances with essences that are not physical or biological but rather mental. Take, for instance, the traditional idea of the soul as an immaterial substance. This idea is commonly placed in a substance dualist (rather than property dualist) framework; but the point that the substance's existence and identity remains unaffected by non-essential change holds analogously: the soul substance might or might not be connected to a body and yet is what it is, the self-same substance.

The assumed priority of the substance over its properties makes the latter appear negligible – comparable to an optional software – unless they have been built into the substance's essence – the hardware, so to speak – via a respective substance concept.¹² Thus, once the decision is made that certain types of properties do not sit at the heart of the substance in question, we can hold fixed the substance's identity without them. The exclusion of subjectivity from the animalist concept of an organism is produced by such an

ontological operation that holds fixed a supposedly biological substance as a self-identical substratum for the coming and going of fleeting and peripheral properties, among which can be, but need not be, psychological ones.

4 Subjectivity beyond personhood: recent empirical discoveries

Is Olson right about the nature of organisms and about their casual relationship with mental states? It is worth asking science about this, in particular those sciences that are concerned with the entities at issue, organisms and the mind.

Philosophers of mind have traditionally been concerned with the *human* mind. They have investigated how the human mind works and how it relates to the body; since the 20th century, in particular, how it relates to the brain. The anthropocentric focus of investigation has been flanked, for a long time, by an utterly intellectualist concept of the mind that emphasises rationality, reflection and conceptual self-consciousness. This intellectualist concept of the mind feeds into the concept of a person employed in the metaphysical debate on personal identity. “[A] person”, Olson (1997, 103) writes with reference to Locke, “is a rational being capable of thought, consciousness, and a certain kind of self-awareness that involves thinking of itself as tracing a path through time and space”. Also following Locke, Olson includes “being a moral agent” in the “list of person-making features”, taking it to be a psychological property too or, at least, to be entailed by psychological properties (Olson 1997, 104). Being or not being a person comes down to possessing or not possessing all these highly sophisticated mental skills found exclusively in humans among natural beings and possibly also in some non-natural beings, such as computers (as suggested by the computational theory of mind) or gods.

In recent years, however, the scope of investigation has broadened. Appreciating that the human mind is an evolved biological function, biologists, cognitive scientists and some philosophers of mind have started to put the human mind into context – the biological context of other natural minds: animal minds (Ginsburg and Jablonka 2019; Godfrey-Smith 2016, 2020; Tye 2017), possibly even plant minds (Calvo 2017; Calvo et al. 2020; Raja and Segundo-Ortín 2021). Research on the evolution of consciousness and cognition is burgeoning, with scholars generally agreeing that mental capacities have evolved gradually and are widespread across species as they exist today. Importantly, there is an emerging consensus that cognitive functions do not require possession of a brain, possibly not even of a nervous system (Baluška and Levin 2016; Smith-Ferguson and Beekman 2020).¹³ Organisms as simple as bacteria have been argued to display behaviour appropriately described in cognitive terms (Lyon 2015; Shapiro 2020). Of course, the cognitive functions of such primitive creatures cannot compete with those of humans. Yet, it turns out there are underlying basic biological mechanisms shared by both. According to the minimal, or basal, cognition approach (Brancazio et al.

2020; Lyon et al. 2021), we ought to understand these basic features of cognition first before endeavouring to unlock the secrets of its more sophisticated forms. This means to analyse cognition into different aspects and capacities, and to trace their occurrences, combinations and variations along the continuum from the bottom up (Lyon et al. 2021; Smith-Ferguson and Beekman 2020).

Given the diversity of cognitive phenomena, one may wonder about the possibility of subsuming all of these under a single, unified concept of cognition. Lyon et al. (2021, 4) offer the following universal definition that is meant to emphasise “the adaptive value” of cognition:

Cognition comprises the sensory and other information-processing mechanisms an organism has for becoming familiar with, valuing, and interacting productively with features of its environment [exploring, exploiting, evading] in order to meet existential needs, the most basic of which are survival/persistence, growth/thriving, and reproduction.

(Lyon et al. 2021, 4)

According to Lyon et al. (*ibid.*), cognition thus understood “is a function necessary for *any* autonomous biological system’s survival, wellbeing and reproduction” (*italics in the original*). This corresponds with Maturana and Varela’s (1980) thesis that all living systems, by being autopoietic, exhibit cognition: the process of self-production (‘auto-poiesis’) involves, and is dependent on, the process of interacting with the environment in an intelligent and, i.e., at least implicitly knowledgeable way.¹⁴ Cognition is embodied and enacted.¹⁵

Attributing (at least) basal cognitive skills to all kinds of organisms does not settle the question of which kinds of organisms are sentient, i.e., have subjective conscious experiences (also known, and controversially debated, as qualia). What is it like to be a bacterium? It is possible that the question is empty because it is not like anything at all to be a bacterium. Bacteria and other prokaryotes might not be subjects of experience. Yet, they could be cognizers. Views among scientists vary on when and how consciousness evolved, and in which extant species it is present. The assessment depends on the respective criteria regarded as necessary and sufficient for consciousness, which in turn depend on what one thinks consciousness is. Ginsburg and Jablonka (2019, 7), for instance, understand consciousness or

subjective experience as a mode of being that involves activities that generate temporally persistent, dynamic, integrated, and embodied neurophysiological states that ascribe values to complex stimuli emanating from the external world, from the body, and from bodily actions.

On the basis of seven corresponding criteria for minimal consciousness,¹⁶ Ginsburg and Jablonka (2019, Chapter 8) argue that apart from vertebrates also arthropods (insects, arachnids, myriapods, crustaceans) and cephalopods

(squid, octopus, cuttlefish, nautilus) fulfil all of these criteria and, therefore, are conscious, this being marked by a shared ability of unlimited associative learning.

The new empirical work on the evolution and distribution of cognition and consciousness demonstrates that organisms that do not qualify as persons, and do not anyhow come near to qualifying as such, still can be subjects of experience. Mental states are not the prerogative of humans and very few other species. Moreover, it becomes evident that characterising cognitive skills as mental properties instantiated, with varying frequency, by different kinds of organism grossly misses the nature of cognition in organisms. Cognition, and at least for many organisms consciousness too, is not merely a non-essential add-on to organic life, something that is nice to have but not strictly required. On the contrary, life is ‘minded’ at its roots: cognition and/or consciousness shapes the living organisation as such (see also Thompson 2010). Ginsburg and Jablonka (2019, see also 2020) indicate this by speaking of consciousness as a ‘mode of being’. Conscious organisms *live differently* from non-conscious ones: their behaviour is to be explained not simply through function but also, and most importantly, through *motivation*:

Consciousness, we suggest, [...] is neither a property nor a process – it is a new teleological, intrinsic mode of being, its teloi being the ascription of values to encountered objects or states through ontogenetically constructed desires that the animal strives to fulfil.

(Ginsburg and Jablonka 2019, 188)

It looks indeed as if there were a fundamental connection between subjectivity and what is sometimes called *valence*: “the biological impetus of attraction or repulsion to a state of affairs based on an assessment of value relative to an individual’s goal structure” (Lyon and Kuchling 2021, 2). Valence, together with arousal, is considered a constituent of affect and, thereby, of emotion, which is “the body’s way of driving the organism to secure its survival, thriving and reproduction” (Lyon and Kuchling 2021, 1 (abstract)). Importantly, there is empirical evidence that affective responses to valenced stimuli do not require consciousness (Lyon and Kuchling 2021, 5f.). This, *pace* Ginsburg and Jablonka, lends support the view that already the behaviour of non-conscious organisms can be explained through motivation (Lyon and Kuchling 2021, 7) – something, in fact, to be expected given the fundamental role valence plays for survival:

Valence arguably is the fulcrum around which the dance of life revolves. Without the ability to discriminate advantage from harm, and to induce behaviour to meet the opportunities and challenges presented by changing circumstances (external and internal), the complex, entropy-dissipating, autopoietic form of organization known as life very quickly comes to an end.

(Lyon and Kuchling 2021, 7; see also 1 (abstract))

If it is true that “valence emerges from the conditions of life as a necessary feature of biological existence” (Lyon and Kuchling 2021, 12) and, per its affective nature, “ineluctably produces a ‘feeling’ that can never be fully described but only experienced” (Lyon and Kuchling 2021, 11), we can further infer that the notion of subjectivity is applicable more broadly, i.e., even to organisms that fail to qualify as subjects of experience in the standard sense. Alternatively, of course, we may want to consider revising the concept of a ‘subject of experience’ by including (hypothetical) non-conscious experiences (possibly of a purely affective kind). Either way, it turns out that cognition as such, however minimally present, might be sufficient for living systems to be subjects – biological subjects.

5 The reappearance of subjectivity in processual animalism

It’s time to back up the empirical findings with some metaphysics.¹⁷ Elsewhere I have argued that to understand organisms as autopoietic systems is to understand them, metaphysically speaking, as processes (Meincke 2019a). I define a process as an entity for whose identity change is essential. Compare Maturana and Varela’s (1980, 78f.) description of an autopoietic system as being

organized (defined as unity) as a network of processes of production (transformation and destruction) of components that produces the components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it [...] as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network.

The network of processes of production of components produces components, which in turn produce the network of processes of production of components. Maturana and Varela (1980, 79f.) are remarkably clear about the fact that nothing in an autopoietic system is static, not even the diachronic relations between the components, i.e., between the parts of the organism. The relations’ relative stability rather results from constant efforts of stabilisation, such as the interactions between the parts and – a point stressed by the more recent theory of biological autonomy (Moreno and Mossio 2015) – the interactions with the environment of the system as a whole. “Organisms”, Varela (1997, 73) declares, “are fundamentally the process of constitution of an identity”, with ‘identity’ “not meant as a static structural description (it is a process)”.

Unlike substances or things, organisms, in other words, cannot just sit there and do nothing; as a minimum requirement of change, metabolism, the process of exchange of matter and energy with the environment by which an

organism rebuilds and renews itself, needs to continue. If it doesn't continue, death will be the consequence. Metabolism is a lower-level process (probably the most fundamental one), which, together with myriads of other lower-level processes, contributes to the constitution of the whole organism as a higher-order process while being itself dependent on the latter. Organisms are therefore best described as orchestrated, hierarchically organised complexes of processes that are mutually dependent and stabilise themselves at different time scales through continuous interaction among themselves and with surrounding processes (Meincke 2018b, 2019c, 2020a).

Conceiving of organisms not as substances but as processes – as interactively self-stabilising processes, according to my suggestion¹⁸ – has implications for the understanding of personhood and personal identity, assuming (as we should) that, at least human, persons are organisms. Under the title of 'processual animalism' I have started developing a metaphysical theory of ourselves that acknowledges both our biological nature and our processual nature. Human persons, qua organisms, are a particular kind of living process. The guiding idea is that whatever distinguishes this kind of living process from others (see Meincke 2019c for some suggestions) is founded upon general bio-processual dynamics and structures, which therefore have to be investigated first. This is not to say that special features of human-kind processes, such as personhood, were *reducible* to biological processes in the sense of being nothing over and above these. Such claims of reductionism neglect the emergent status of personhood and the fact that personhood-related dynamics feed back causally on biological dynamics. Most importantly, however, we ought to realise that biology itself comprises much more than modern mechanistic and, indeed, reductionist thinking has been prepared to accept or even to imagine; that biology might even be different in a fundamental way altogether. Appreciating the processual nature of organisms facilitates such a more inclusive notion of 'biology'.

Processual animalism, as far as the biological basis is concerned, takes seriously the processual nature of organisms in three fundamental respects.

First, as explained elsewhere in more detail (Meincke 2018b, 2019a, 2020a), *biological identity* (see also Meincke and Dupré 2020a) is reconceptualised as a process and, that is, as an achievement rather than a primitive pre-given fact. In direct contrast to substances, which even when undergoing change retain an unchanging core, i.e., their essence, change is all-pervasive in organisms; and, furthermore, rather than being preceded by an identity which it either does not affect or actually jeopardises, change, in organisms, is what *constitutes* identity (provided, of course, it is of the right kind and occurs at the right time). An organism persists exactly as long as it keeps changing aptly and, that is, as long as it manages to stabilise itself through appropriate interactions, which necessarily includes performing ex-changes of matter and energy with the environment.¹⁹

Second, processual animalism entails a process account of *bio-agency*.²⁰ The central insight here is that actions, such as instances of sensorimotor

behaviour, are a particular, namely, more sophisticated form of the interactions with the environment which every organism has to perform in order to survive, i.e., persist. Whether a tiger hunts for a prey or digests it, either is needed for survival. In fact, action – e.g., hunting for prey – is what keeps digestion and, hence, metabolism, going for tigers. Agency, thus, cannot be understood independently of metabolism, which could be seen as a minimal form of agency itself and from which more robust forms of agency appear to have evolved through a stepwise emancipation (Meincke 2018a; Moreno and Etxeberria 2005; Moreno and Mossio 2015).²¹ To the extent that, according to processual animalism, both bio-agents and their actions are processes,²² actions are not merely extrinsic features of agents, changes they accidentally undergo and that do not go deep. Instead, whatever an organism does is part of the interactive process of self-stabilisation that is the organism. Biological identity and bio-agency are thus ontologically entangled.²³ Bio-agency is a condition of biological identity for bio-agents; and biological identity facilitates bio-agency. Biological identity, qua processual, is (at least minimally) agential identity.

On the basis of the process accounts of biological identity and of bio-agency put forth by processual animalism, we can now turn to its third element: a process account of *biological subjectivity*. In the previous section we learned that a growing body of empirical studies of cognition and consciousness in non-human organisms points towards the existence of a form of subjectivity that does not require any intellectual skills, let alone personhood. Instead, organisms appear to be subjects simply as a result of the way they are organised as living systems, which arguably involves (at least minimal) cognition and thereby some form of experiencing. Where these experiences indeed happen below the threshold of consciousness, subjectivity is still indicated by the organism's motivation to respond to valenced stimuli, namely by performing an appropriate action.²⁴ In other words, *agency* is the context in which valence, affective experience and subjectivity are located.

The connection between subjectivity and agency has not gone unnoticed. Thus, Godfrey-Smith (2020, 105) explains:

As evolution proceeded, animals became a new kind of intersection point, or nexus, in the world's networks of causal pathways. [...] Animals of this kind have a point of view, and from that viewpoint they act. Some of what falls under the problematic umbrella of 'subjective experience' is an expected, comprehensible consequence of animal evolution. Roughly speaking, the evolution of animal agency brings with it the origin of subjects.

Actions are perspectival just like the values to which they respond; actions are a means to obtaining or realising what a particular agent in a particular situation within a particular environment deems worth striving for. According to Godfrey-Smith, this entails some sort of dichotomy between self

and world. Subjective experience is the “the shaping of living activity into a *self* with a perspective” (Godfrey-Smith 2020, 222; italics in the original); “[s]ubjectivity is about engaging with the world as a self, and having things seem a certain way *to you*” (Godfrey-Smith 2020, 223; italics in the original).

How can we account for this connection between agency and subjectivity, and its implication of selfhood, in metaphysical terms? It is very hard to provide such an account within a substance ontological framework. Admittedly, the history of western metaphysics is replete with substantial selves – from Descartes’s *res cogitans* over Leibniz’s monads to E. J. Lowe’s simple, non-composite objects of first-person reference or Richard Swinburne’s individual souls.²⁵ But apart from the fact that these views invoke highly intellectualist conceptions of the mind, none actually delivers an explanation as to why the selves of the supposed kind exist and where they have their characteristic intellectual capacities from.

This is no accident, given that substances are believed to be ontologically independent entities. Their identity, as already indicated, does not depend on anything but is a primitive. Accordingly, a substance’s identity also does not depend on the substance’s actions. If a substance indeed acts, this is not because it *has to* in order to persist. Instead, actions are taken to flow from the substance’s essence, as an expression of its inner, pre-given nature. This implies their being only contingently related to external circumstances. A substance’s environment does not play a constitutive role for the actions of the substance, let alone for the latter’s existence. As it happens, the very idea of an environment, if considered at all, has traditionally been under suspicion. Descartes famously needed to appeal to God’s existence in order to prove the existence of an outer material world in addition to the undoubtable existence of the thinking substance, and Leibniz held that bodies exist only as the objects of perceptions of monads.

Within a substance ontological framework, agency is not on the radar as a possible source of subjectivity because it is regarded as parasitic upon, rather constitutive of, identity. Selfhood is posited, not explained. The much more attractive alternative consists in adopting a process ontological framework and explaining subjectivity in terms of a dynamical agential self-world relationship.

Organisms need to survive. From a process ontological perspective, this fact is anything but banal. It means that an organised complex of processes must interact with other processes in a manner that makes sure that itself and these other processes, though interactively intertwined, yet do not collapse into one another. This basic situation of living systems is as precarious as paradoxical. Precarious because the collapse, as the second law of thermodynamics tells us, is inevitable and can only be delayed, namely for exactly as long as a living system, qua thermodynamically open system, manages to uphold a far-from-equilibrium steady state by entertaining a constant exchange of matter and energy with the environment. Paradoxical because exactly in order to avoid collapsing into the surrounding processes, a living process needs

to engage with these; an organism's temporary, fragile identity only exists, as it were, on the back of what it is not identical with, the environment which it exploits for the satisfaction of its needs.

What I want to propose here is that this dynamical, dialectical relationship between identity and non-identity, built into the processual constitution of organisms as interactively self-stabilising higher-order processes, is, meta-physically speaking, the origin of subjectivity. Living systems exist through their own efforts of self-demarcation from surrounding processes, on which they, at the same time, depend and which, therefore, are meaningful to them: the environment as a space of potential satisfaction of needs co-emerges with 'needy' living processes. This co-emergence is tantamount to the co-emergence of self and non-self, subject and object, and explains the observed nexus between affective experience, agency and valence. Containing the objects of want of a subject, appropriated by the latter through affection and action, the environment presents itself to that subject as a treasury of value. Subjectivity, most fundamentally, just *is* the process of valencing possibilities for actions, enabled by, and inherent in, the process of establishing and maintaining a self/non-self-distinction through the interactive construction of a boundary between organism and environment.

The concept of a boundary plays a crucial role in the autopoiesis theory of life. According to Varela (1981, 14f.), "the idea of autopoiesis is, by definition, restricted to relations of production of some kind, and refers to topological boundaries". This is expressed in the definition of an autopoietic system quoted above through the requirement that the components of a network of production of components 'specify the topological domain' of the network's realisation. Autopoietic, or living, boundaries ought to be understood correctly, though.²⁶ Most importantly, one must notice that living boundaries do not completely isolate the living system from its environment; otherwise the interaction with the environment, which is constitutive of the self/non-self-distinction, would be impossible. Living boundaries, therefore, contrast sharply with the boundaries of a substance. Instead of being strict, neat and parasitic upon a pre-given identity, they are processually generated, flexible and semipermeable: their function is to facilitate a selective and controlled interaction with the environment.²⁷ Not just everything is allowed in but only that which contributes to the organism's survival and well-being. As we have seen, this selective and controlled character of organism-environment interaction is associated, in both the autopoiesis theory and today's debates on minimal cognition, with the presence of cognition.

Hans Jonas (1968, 2001 [1966]) has powerfully described the precarious and paradoxical situation of life in terms of a 'needful freedom' of form from matter ensuing from the metabolic constitution of living systems (see Meincke 2018b). In metabolism, form emancipates itself from matter by becoming the cause of the flow of matter through it, while at the same time remaining dependent on matter insofar as there would be no form without matter arranged such and such over time. Naturally, then, living form is both

“truly pitted against the rest of things” by defying “the equalizing forces of physical sameness all around” (Jonas 2001, 83) and “turned outward and toward the world”, entertaining with it a “vital relationship” (Jonas 2001, 84).

According to Jonas, this dialectical relationship between ‘self-isolation’ from and ‘transcendence’ towards the world is at the roots of the phenomenon of selfhood, characterising the ‘internal identity’ of the living as opposed to the ‘external identity’ of non-living material things and necessarily incorporating subjective experience.²⁸ The “essential boundary dividing ‘inside’ and ‘outside’” (Jonas 2001, 79), which constitutes the organism’s “[p]rofound singleness and heterogeneity within a universe of homogeneously interrelated existence” (Jonas 2001, 83), therefore paradoxically becomes the enabling condition of an even more intense and intimate relationship than to be found among non-living material things, which, as Jonas (1968, 2001, 79) argues, lack genuine ontological individuality. This self-world relationship is mediated through agency as an affection-driven, experience-informed mode of selective interaction with the world:

The primary solicitation of transcendence indeed issues from the organic want and is thus one with the commitment to activity: it is outgoing; but receptivity of sense for the incoming, this passive side of the same transcendence, enables life to be selective and ‘informed’ instead of a blind dynamism. Thus, in facing outward, internal identity becomes the subject-pole of a *communication with things more intimate than that between mere physical units*, and so the *very opposite of isolation* emerges from the isolation of the organic self.

(Jonas 2001, 85; italics added)

Subjectivity is communication, an openness towards the other, in which, as an object, a subject takes interest as part of the interest it takes in itself. No conceptual skills, no reflection, no intellect, possibly not even consciousness is required for this.²⁹ Living organisation and, that is, interactive processual constitution of a boundary that both separates from and connects with what is beyond it is enough.

6 Conclusions

By arguing that ‘our’ identity through time depends on purely biological conditions, namely on relations of biological continuity, Olson asks us to accept that the ontological explanation of numerical identity provided by animalism cannot account for the prudential, moral and social aspects of ‘our’ identity. Accounting for these, in a wide sense, normative aspects is a task to be delegated to a complementary theory that tracks relations of psychological continuity through time. Alas, in hypothetical cases where biological continuity and psychological continuity go different ways, animalist ontology and psychological ethics, we have seen, yield contradictory identity judgements,

which raises doubts about animalism's claim to ontological priority over psychological accounts.

This difficulty is exacerbated by Olson's extensive references to human animals with the first-person pronouns 'I' and 'we', suggesting subjectivity even when, according to his own theory, there is none because no psychological states are present. Indeed, it is exactly a general exclusion of subjectivity from biology that causes the dissociation of the ontology and the ethics of personal identity. Metaphysically, organisms are conceptualised as substances with purely biological essences in the sense that no psychological properties contribute to the organism's identity. Psychological properties are mere accidents and, at the same time, strictly distinguished from biological properties.

As it happens, Olson's substance-ontological-cum-property-dualist view of the organism appears to be wrong – both empirically and metaphysically. Recent empirical research on the evolution of the mind and on minimal cognition testifies to a global applicability of the notion of a subject of experience in the animal kingdom and possibly beyond. It's not only that subjectivity must not be equated with personhood; subjectivity rather appears to be inextricably intertwined with the organisation of living systems as such. Thus, even where we might be in doubt about the conscious quality of experience, selective responses to valenced stimuli indicate at least basic cognition and a rudimentary subjective perspective.

A process ontological framework backs up the empirical findings. If organisms are not ontologically independent particulars whose identity is primitively given through an immutable essence, but if they are rather dynamical entities whose identity is the effect of their own continuous efforts, we can understand how subjectivity emerges as a biological function: having continuously to stabilise itself as a complex higher-order process through interaction with surrounding processes, a living system is in itself related, in a characteristic combination of opposition and need, towards what is beyond its arduously maintained boundary. The co-existence of self and world, subject and object, reflects the processual constitution of living boundaries, which enable interactions with the environment as much as they are themselves the result of such interactions. Having to interact in order to persist entails having a perspective.³⁰

Organisms, human and non-human, are not substances with purely biological essences devoid of psychology; they are no substances at all. Organisms are processes, namely processes which, qua biological, are inherently subjective.³¹ This integrative understanding of 'biological', which, together with processual reconceptualisations of biological identity and agency, marks the transition from substance ontological animalism to *processual animalism*, impacts on animalism's relationship with ethics. Simply put: as it turns out that we cannot tear apart biology and psychology, there is no need to tear apart the ontology and the ethics of personal identity either.

There are two aspects to this. First, there is no reason to confine the relevance of psychology to ethics. The possession of mental lives, however

minimal they may be, is an ontological feature of animals or even of organisms generally. Psychology, therefore, has the same claim to ontological authority as biology has. Second, biology matters ethically when it comes to questions of personal identity. Olson's animalism concedes unnecessarily too much to the Psychological Approach when he ties prudential, moral and social concerns about our identity exclusively to psychological continuity on the basis of an over-intellectualised view of the mind. Organisms, constantly threatened with annihilation, are subject to biological norms of success, failure and well-being (Burge 2009); and it is simply not true that we only care about whether our more sophisticated, reflective and rational skills live on. What matters to us in caring about our, or our loved ones', identity is the whole orchestrated living process that strives for survival within its environment, including its perceptions, feelings, desires, and its joy and suffering.

That there is a perspectival or subjective³² element in what is otherwise presented by Olson as psychology-free biology, and that this perspectival element does matter somehow, is evinced by the occurrence of first-person subjective references in unexpected contexts. What would be the point of calling animalism a theory of 'our' identity if the human animals 'we' are failed to be *us* in a not merely grammatical sense? What could it mean to say that 'I' once was a foetus and that 'I' one day may be in a PVS if there was no ontological ground whatsoever to relate to the foetus and the PVS patient? Processual animalism allows us to recognise ourselves in the foetus and the PVS patient as the same living subject, the former being on track towards becoming a fully fledged subject of experience in accordance with the typical life cycle of human organisms, the latter being in a deplorable state of having been deprived of their higher cognitive functions including access to their own experiences. Processual animalism, in other words, legitimises using subjectivity-laden language when talking about foetal and PVS phases of human animal life.

Parfit (1987) famously argued that 'what matters' is not numerical identity but survival. This seems roughly right; however, not in the sense intended by Parfit and conceded to him in an ethical reinterpretation by Olson. Survival in biology is in fact rather different from Parfitian survival. It is, in particular, in no way comparable to the unhindered, potentially endless running of some (psychological) software on some possibly changing (physical) hardware, while possibly, and potentially endlessly, being reduplicated. I have described here, in metaphysical terms, survival for organisms as a matter of interactive self-stabilisation. This fundamental ontological process of life can be realised in a stunning variety of ways, illustrated by the biodiversity on Earth. As different species adapt to different environments in different ways, the concrete, biological requirements for survival vary considerably. In this sense, Olson's (1997, 30) claim that "our persistence conditions are the same as those of aardvarks and oysters and other animals" clearly does not hold.

Human lives, as indicated, unfold in life cycles, moving from gestation to infancy, childhood, adolescence, adulthood, old age and, eventually, death. The changes involved in this process encompass the organism as a whole – its

brain as much as the rest of the body. Additionally, humans develop a personality through social processes of interactive self-stabilisation, constituting their ‘personal identity’ in a more emphatic and specific sense (Meincke 2019c). As I intend to elaborate in future work, our survival as the persons we are, with our individual, historically developed personalities, depends on our relationships with other persons and with the environment more generally: hostile social environments may kill us; hospitable social environments let us flourish.

Within this framework set out by processual animalism, hypothetical cerebrum transplantations are to be regarded as interferences with some, admittedly important, aspects of what it means to be a human animal. Olson is right that in such cases one animal loses an organ while another animal gains a new organ. However, I disagree with Olson’s assumption that there is a sense in which cerebrum transplants move a person from one place to another, an assumption that opens the door for the unfortunate dissociation of ontological and ethical assessments of our identity through time. A human person does not sit in the cerebrum like a homunculus; a human person is rather embodied in the whole of the organism including in its past and present interactions with the environment.³³ Personhood is holistic and historical. It is therefore not clear that loss of cerebral functions is tantamount to loss of personhood. At any rate, a hypothetical cerebrum transplantation can be expected, among other effects, to remove access to one’s own basic experiences similarly as lapsing into a PVS state does; but as we have seen, this is compatible with the continued presence of a subject of experience.³⁴ Common practices of ethical concern for PVS patients reflect this, too.³⁵

Animalism, I conclude, does not have to give up on guaranteeing “the coincidence of what is important in our identity with the actual conditions of our identity” (Olson 1999, 165). If we take seriously the processual organisation of organisms in general and of human organisms in particular, the supposed dualism between biological properties on the one hand and psychological properties on the other vanishes, this in accordance with the empirical findings in biology and cognitive science. Biology and subjectivity are not enemies but friends – and so are the ontology and the ethics of personal identity.

Acknowledgements

This chapter has been made possible by an Elise Richter research grant from the Austrian Science Fund (Fonds zur Förderung der wissenschaftlichen Forschung/FWF), grant agreement number V 714. It is based on a talk given at the conference of the *pontology* network “The Unity of a Person: Ontology – Subjectivity – Normativity” in Munich on April 26, 2019. I wish to thank the organiser, Jörg Noller, for the invitation and the audience for stimulating discussion. I am also grateful to Menno Lievers for helpful comments on the first draft of this chapter.

Notes

- 1 Olson (2015) wants to reserve the term ‘animalism’ for the claim that we are animals, while speaking of ‘the Biological Approach’ with respect to our identity through time insofar as the conditions of our identity of time are taken to be purely biological. The latter, he argues, follows from what he calls animalism only together with independent assumptions about the nature of animals. Here I shall ignore this complication and use the terms ‘animalism’ and ‘Biological Approach’ interchangeably.
- 2 Not everyone understands animalism this way. Wiggins (2001), for example, is often seen as presenting an animalist position despite the fact that his account incorporates psychological aspects. Snowdon (2014) agrees with Olson that no psychological states are required for our existence, but rejects, among other things, Olson’s (and van Inwagen’s) assumption that animals cease to exist at death.
- 3 For other worries, see Meincke (2020a) and endnote 6.
- 4 The theories Olson refers to as ‘the Psychological Approach’ belong to the family of so-called Complex or Reductionist theories of personal identity, which treat the latter as a complex phenomenon that can be reduced to empirical relations, such as relations of continuity. In contrast, so-called Simple or Non-reductionist theories of personal identity deny any reducibility, regarding personal identity instead as a simple phenomenon, a ‘deep further fact’ adding to the empirical facts about persons. (Swinburne 1984, for instance, appeals to the identity of an immaterial individual soul.) Olson (1997, 4) presents animalism as a Complex View; see, however, endnote 6.
- 5 Baker (1999, 152ff.) accuses Olson of misconstruing the Psychological Approach as being committed to the stronger *de re* thesis ‘If *x* is a person, then *x* necessarily has psychological properties’ rather than to the weaker *de dicto* thesis ‘Necessarily, if *x* is a person, then *x* has psychological properties’. If this is true and the Psychological Approach is committed only to the latter thesis, then Olson’s conclusion would not follow, i.e., the Psychological Approach would not need to deny that a person once was a foetus and may be a PVS patient one day.
- 6 This supposition is wrong, as I have argued in Meincke (2020a). A Biological Approach to personal identity that relies on biological continuity faces the same problems of fission and fusion in the form of real-world scenarios which the Psychological Approach faces (as of yet) as mere science-fiction scenarios figuring in fancy thought experiments. Where animalism tries to evade these problems, e.g., when appealing to a Life Criterion of identity (Olson 1997, 135–140; see also van Inwagen 1990), it lapses into a Simple View, i.e., becomes circular and uninformative, see Meincke (2015, 125–143, 2020a).
- 7 Things are obviously more complicated in the case of whole-brain transplantations: “Removing an animal’s entire brain either kills the animal or reduces it to a mere detached brain” (Olson 1997, 45). This is why Olson (1997, 44–46) insists on distinguishing between cerebrum transplantations and whole-brain transplantations as well as between the respective intuitions.
- 8 See endnote 6.
- 9 This analogy is not meant to invoke any functionalist views of the mind.
- 10 See Meincke (2010) for a detailed discussion of Olson’s problematic use of subjective language.
- 11 Olson seems to assume that mental properties supervene upon physical properties (see Olson 2007, 34); and this supervenience relation would still hold in those cases where numerical identity (ontology) and ‘personal identity’ (ethics) come apart as a result of cerebrum transplantation. At the same time, however, Olson does not want to rule out the existence of non-human persons, such as “intelligent Martians, gods, angels, demons, trolls, or even rational, conscious

electronic computers made of metal and silicon” (Olson 1997, 17, see also 30 and 142). Olson does not say whether he would regard the immaterial persons among these as immaterial substances. The mental properties of immaterial substances would not supervene upon physical properties.

- 12 Accordingly, in the substratum theory, which, unlike the substance theory, does without the assumption of essential properties, *all* properties attributed to the substratum are considered irrelevant to the substratum’s identity.
- 13 This amounts to a radicalisation of the embodied cognition approach in cognitive science, according to which cognition is not confined to the brain.
- 14 See Maturana and Varela (1980, 13):

A cognitive system is a system whose organization defines a domain of interaction in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in the domain. Living systems are cognitive systems, and living as a process is a process of cognition. This statement is valid for all organisms, with or without a nervous system.

- 15 “Cognition is not the representation of a pre-given world by a pre-given mind but is rather the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs” (Varela et al. 1991, 9). See also Noë (2004) and Gallagher (2017).
- 16 These are: (1) global activity and differentiated states; (2) binding/fusion and unification; (3) intentionality; (4) selection, plasticity, learning and attentional modulation; (5) temporal thickness; (6) attribution of values, emotions, goals; (7) embodiment and self (Ginsburg and Jablonka 2019, especially 98ff.).
- 17 I take it that metaphysics ought to be informed by science just as much as philosophers of science (and, to some extent, scientists) benefit from familiarity with metaphysical concepts and arguments. See Meincke and Dupré (2020b) for a programmatic motivation of such cross-disciplinary dialogue. See Meincke (forthcoming) for a more detailed ontological clarification of the notion of process.
- 18 See Nicholson and Dupré (2018) for other versions of a process (philosophy of) biology.
- 19 Biological identity, understood as a process of interactive self-stabilisation, is weaker than substance identity but stronger than mere biological continuity. Thus, cases of fission and fusion can be accommodated in a convincing way (Meincke 2020a). For arguments in favour of a general process account of persistence, see Meincke (2019b).
- 20 Working out the details of this account is part of the agenda of my current research project “Bio-Agency and Natural Freedom”, funded by the Austrian Science Fund (FWF).
- 21 The perhaps most important step in the course of this emancipation, the evolution of functional body movement in an environment, was facilitated by the evolution of the nervous system (Arendt 2021; Jékely et al. 2021; Moreno and Mossio 2015).
- 22 In the context of analytic metaphysics, Steward (2012) was the first to claim that actions are processes rather than (as the orthodox view has it) events. However, Steward conceptualises agents as substances.
- 23 Therefore, contrary to traditional practice in analytic metaphysics, they cannot be examined in isolation.
- 24 Note that reflex reactions lie on a continuum with (‘proper’) actions and do not rule out motivation and affection.
- 25 See, e.g., Lowe (1991) and Swinburne (1984).
- 26 I need to defer an in-depth discussion of living boundaries to a separate paper. For some reflections on boundaries in the context of the autopoiesis theory of life, see Meincke (2020b).

- 27 This is paradigmatically true for biological membranes but also for non-topological living boundaries, such as the immune system.
- 28 “[...] there is inwardness or subjectivity involved in this transcendence [of organic form towards the world, A. S. M.], imbuing all the encounters occasioned in its horizon with the quality of felt selfhood, however faint its voice. It must be there for satisfaction or frustration to make a difference. Whether we call this inwardness feeling, sensitivity and response to stimulus, appetite or *nîsus* – in some (even if infinitesimal) degree of ‘awareness’ it harbors the supreme concern of organism with its own being and continuation of being – that is, it is self-centered – and at the same time bridges the qualitative gulf to the rest of things by selective modes of *relation* which, with their specificity and urgency, replace for the organism the general integration of material objects in their physical context” (Jonas 2001, 84; italics in the original).
- 29 Evolutionary later, more developed forms of living organisation show, of course, a richer phenomenology of selfhood. Jonas (2001, 99–107) identifies as a major step the evolution of the nervous centralisation of animal life, which he regards as coincident with the evolution of sentience and motility (see also Jonas 1968, 244f.). For an empirically founded study on selfhood in plants, see Baluška and Mancuso (2021). Zahavi (2005) defends a minimal notion of selfhood from a phenomenological perspective; see also Schlicht (2018). In contrast, Stapleton and Froese (2016) insist that subjectivity requires the self to be an explicit part of an organism’s experience, involving a reflective stance.
- 30 Susan Hurley refers to the corresponding aspects of self-consciousness as “perspectival consciousness”: “Part of what it is to be in conscious states, including perceptual states, is to have a unified perspective, from which what you perceive depends systematically on what you do and vice versa, and such that you keep track, at the personal level, of this interdependence of perception and action” (Hurley 2003, 234). Jékely et al. (2021) speak of ‘body-selves’ whenever action and sensing are linked by reafference.
- 31 This implies that subjectivity, for organisms, is not just a psychological property instantiated by a biological substance among other psychological properties. Subjectivity, like consciousness, is (to use Ginsburg and Jablonka’s 2019 term) a – or rather: the – ‘way of life’ for organisms.
- 32 I agree with Godfrey-Smith (2020, 299) that “all points of view are first-person”.
- 33 This explains, for instance, the possibility that memories lost through dementia can (at least partly) be restored through supportive social interactions.
- 34 In fact, recent research provides evidence even for phenomenal consciousness in PVS patients. For an empirically informed philosophical discussion of PVS, see Shea and Bayne (2010). See also Solms (2021) for a state-of-the-art discussion of the neuroscience of consciousness.
- 35 Likewise, we can exclude the possibility left open by Olson (1997, 45) that whole-brain transplantations amount to shrinking an animal to the size of a brain and moving it about (see endnote 7). A brain in a vat does not interact with the environment as an animal does; there is no interactive self-stabilisation – no metabolism, no agency. See also Thompson and Cosmelli (2011).

Bibliography

- Arendt, Detlev (2021), “Elementary Nervous Systems”, in: *Philosophical Transactions of the Royal Society B* 376, 20200347.
- Baker, Lynne Rudder (1999), “What Am I?”, in: *Philosophy and Phenomenological Research* 59(1), 151–159.

- Baluška, František and Levin, Michael (2016), “On Having no Head: Cognition throughout Biological Systems”, in: *Frontiers in Psychology* 7, 902.
- Baluška, František and Mancuso, Stefano (2021), “Individuality, Self and Sociality of Vascular Plants”, in: *Philosophical Transactions of the Royal Society B* 376, 20190760.
- Brancazio, Nick, Segundo-Ortín, Miguel and McGivern, Patrick (2020), “Approaching Minimal Cognition: Introduction to the Special Issue”, in: *Adaptive Behavior* 28, 401–405.
- Burge, Tyler (2009), “Primitive Agency and Natural Norms”, in: *Philosophy and Phenomenological Research* 79(2), 251–278.
- Calvo, Paco (2017), “What Is It Like to Be a Plant?”, in: *Journal of Consciousness Studies* 24, 205–227.
- Calvo, Paco, Baluška, František and Trewavas, Tony (2020), “Integrated Information as a Possible Basis for Plant Consciousness”, in: *Biochemical and Biophysical Research Communications*, <https://doi.org/10.1016/j.bbrc.2020.10.022>.
- Gallagher, Shaun (2017), *Enactivist Interventions: Rethinking the Mind*, Oxford.
- Ginsburg, Simona and Jablonka, Eva (2019), *The Evolution of the Sensitive Soul. Learning and the Origins of Consciousness*, Cambridge, MA.
- Ginsburg, Simona and Jablonka, Eva (2020), “Consciousness as a Mode of Being”, in: *Journal of Consciousness Studies* 27, 148–162.
- Godfrey-Smith, Peter (2016), *Other Minds. The Octopus and the Evolution of Intelligent Life*, London.
- Godfrey-Smith, Peter (2020), *Metazoa. Animal Minds and the Birth of Consciousness*, London.
- Hurley, Susan L. (2003), “Animal Action in the Space of Reasons”, in: *Mind and Language* 18, 231–256.
- Jékely, Gáspár, Godfrey-Smith, Peter and Keijzer, Fred (2021), “Reafference and the Origin of the Self in Early Nervous System Evolution”, in: *Philosophical Transactions of the Royal Society B* 376, 20190764.
- Jonas, Hans (1968), “Biological Foundations of Individuality”, in: *International Philosophical Quarterly* 8, 231–251.
- Jonas, Hans, (2001) [1966], *The Phenomenon of Life. Towards a Philosophical Biology*, Evanston, IL.
- Lowe, E. Jonathan (1991), “Substance and Selfhood”, in: *Philosophy* 66, 81–99.
- Lyon, Pamela (2015), “The Cognitive Cell: Bacterial Behavior Reconsidered”, in: *Frontiers in Microbiology* 6, 264.
- Lyon, Pamela and Kuchling, Franz (2021), “Valuing What Happens: A Biogenic Approach to Valence and (Potentially) Affect”, in: *Philosophical Transactions of the Royal Society B* 376, 20190752.
- Lyon, Pamela, Keijzer, Fred, Arendt, Detlev and Levin, Michael (2021), “Reframing Cognition: Getting Down to the Biological Basics”, in: *Philosophical Transactions of the Royal Society B* 376, 20190750.
- Maturana, Humberto R. and Varela, Francisco G. (1980), *Autopoiesis and Cognition: The Realization of the Living*, Dordrecht.
- Meincke, Anne Sophie (2010), “Körper oder Organismus? Eric T. Olsons Cartesianismusvorwurf gegen das Körperkriterium transtemporaler personaler Identität”, in: *Philosophisches Jahrbuch* 117, 88–120.
- Meincke, Anne Sophie (2015), *Auf dem Kampfplatz der Metaphysik. Kritische Studien zur transtemporalen Identität von Personen*, Münster.
- Meincke, Anne Sophie (2018a), “Bio-Agency and the Possibility of Artificial Agents”, in: Christian, Alexander, Hommen, David, Retzlaff, Nina and Schurz,

- Gerhard (eds.), *Philosophy of Science. Between the Natural Sciences, the Social Sciences, and the Humanities*, Cham, 65–93.
- Meincke, Anne Sophie (2018b), “Persons as Biological Processes. A Bio-Processual Way-Out of the Personal Identity Dilemma”, in: Nicholson, Daniel J. and Dupré, John (eds.), *Everything Flows. Towards a Processual Philosophy of Biology*, Oxford, 357–378.
- Meincke, Anne Sophie (2019a), “Autopoiesis, Biological Autonomy, and the Process View of Life”, in: *European Journal for Philosophy of Science* 9, 5.
- Meincke, Anne Sophie (2019b), “The Disappearance of Change: Towards a Process Account of Persistence”, in: *International Journal of Philosophical Studies* 27, 12–30.
- Meincke, Anne Sophie (2019c), “Human Persons – A Process View”, in: Noller, Jörg (ed.), *Was sind und wie existieren Personen? Probleme und Perspektiven der gegenwärtigen Forschung*, Münster, 57–80.
- Meincke, Anne Sophie (2020a), “Processual Animalism”, in: Meincke, Anne Sophie and Dupré, John (eds.), *Biological Identity: Perspectives from Metaphysics and the Philosophy of Biology, (History and Philosophy of Biology)*, London, 251–278.
- Meincke, Anne Sophie (2020b), “Systems or Bodies? On How (Not) to Embody Autopoiesis”, in: *Adaptive Behavior* 28(2), 119–120.
- Meincke, Anne Sophie (forthcoming), “Bergson and Process Philosophy of Biology”, in: Sinclair, Mark and Wolf, Yaron (eds.), *The Bergsonian Mind*, London.
- Meincke, Anne Sophie and Dupré, John (eds.) (2020a), *Biological Identity. Perspectives from Metaphysics and the Philosophy of Biology, (History and Philosophy of Biology)*, London.
- Meincke, Anne Sophie and Dupré, John (2020b), “Biological Identity. Why Metaphysicians and Philosophers of Biology Should Talk to One Another”, in: Meincke, Anne Sophie and Dupré, John (eds.), *Biological Identity. Perspectives from Metaphysics and the Philosophy of Biology, (History and Philosophy of Biology)*, London, 1–21.
- Moreno, Alvaro and Etxeberria, Arantza (2005), “Agency in Natural and Artificial Systems”, in: *Artificial Life* 11, 161–175.
- Moreno, Alvaro and Mossio, Matteo (2015), *Biological Autonomy. A Philosophical and Theoretical Enquiry*, Dordrecht.
- Nicholson, Daniel J. and Dupré, John (eds.) (2018), *Everything Flows. Towards a Processual Philosophy of Biology*, Oxford.
- Noë, Alva (2004), *Action in Perception*, Cambridge, MA.
- Olson, Eric T. (1997), *The Human Animal. Personal Identity Without Psychology*, Oxford.
- Olson, Eric T. (1999), “Reply to Lynne Rudder Baker”, in: *Philosophy and Phenomenological Research* 59, 161–166.
- Olson, Eric T. (2007), *What Are We? A Study in Personal Ontology*, Oxford.
- Olson, Eric T. (2015), “What Does It Mean to Say That We Are Animals?”, in: *Journal of Consciousness Studies* 22, 84–107.
- Parfit, Derek (1987), *Reasons and Persons*, Oxford.
- Raja, Vicente and Miguel, Segundo-Ortín (eds.) (2021), *Plant Sentience: Theoretical and Empirical Issues*, *Journal of Consciousness Studies* 28(1–2).
- Schlicht, Tobias (2018), “Experiencing Organisms: From Mineness to Subject of Experience”, in: *Philosophical Studies* 175, 2447–2474.
- Shapiro, James A. (2020), “All Living Cells Are Cognitive”, in: *Biochemical and Biophysical Research Communications*, <https://doi.org/10.1016/j.bbrc.2020.08.120>.
- Shea, Nicholas and Bayne, Tim (2010), “The Vegetative State and the Science of Consciousness”, in: *British Journal for the Philosophy of Science* 61(3), 459–484.

- Shoemaker, David (2016), "The Stony Metaphysical Heart of Animalism", in: Blatti, Stephan and Snowdon, Paul F. (eds.), *Animalism. New Essays on Persons, Animals, and Identity*, Oxford, 303–327.
- Smith-Ferguson, Jules and Beekman, Madeleine (2020), "Who Needs a Brain? Slime Moulds, Behavioural Ecology and Minimal Cognition", in: *Adaptive Behaviour* 28, 465–478.
- Snowdon, Paul F. (2014), *Persons, Animals, Ourselves*, Oxford: Oxford University Press.
- Solms, Mark (2021), *The Hidden Spring: A Journey to the Source of Consciousness*, London: Profile Books.
- Stapleton, Mog and Froese, Tom (2016): "The Enactive Philosophy of Embodiment: From Biological Foundations of Agency to the Phenomenology of Subjectivity", in: García-Valdecasas, Murillo, José Ignacio and Barrett, Nathaniel F. (eds.), *Biology and Subjectivity. Philosophical Contributions to Non-reductive Neuroscience*, Cham, 113–129.
- Steward, Helen (2012), "Actions as Processes", in: *Philosophical Perspectives* 26, 374–388.
- Swinburne, Richard (1984), "Personal Identity: The Dualist Theory", in: Shoemaker, Sydney and Swinburne, Richard, *Personal Identity*, Oxford: Blackwell, 1–66.
- Thompson, Evan (2010), *Mind in Life. Biology, Phenomenology and the Sciences of Mind*, Harvard.
- Thompson, Evan and Cosmelli, Diego (2011), "Brain in a Vat or Body in a World? Brainbound versus Enactive Views of Experience", in: *Philosophical Topics* 39(1), 163–180.
- Tye, Michael (2017), *Tense Bees and Shell-Shocked Crabs. Are Animals Conscious?*, Oxford.
- Van Inwagen, Peter (1990), *Material Beings*, Ithaca, NY and London.
- Varela, Francisco J. (1981), "Autonomy and Autopoiesis", in: Roth, Gerhard and Schwegler, Helmut (eds.), *Self-organizing Systems: An Interdisciplinary Approach*, Frankfurt, 14–24.
- Varela, Francisco J. (1997), "Patterns of Life: Intertwining Identity and Cognition", in: *Brain and Cognition* 34, 72–87.
- Varela, Francisco J., Rosch, Eleanor and Thompson, Evan (1991), *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, MA.
- Wiggins, David (1980), *Sameness and Substance*, Oxford.
- Wiggins, David (2001), *Sameness and Substance Renewed*, Cambridge.
- Zahavi, Dan (2005), *Subjectivity and Selfhood*, Cambridge, MA.

8 Coherence in the self-pattern

Shaun Gallagher

1 Self-patterns

In contrast to reductionist or deflationist conceptions of the self or person, I want to defend a pluralist view. One can find pluralist views in theorists such as William James and Ulrich Neisser. James (1890), for example, distinguished between the physical self, the social self, and the private self, all of them understood to be part of what it means to be human. Continuing in this tradition, Ulrich Neisser (1988, 1991) discussed five types of self-knowledge corresponding to different aspects of the self, including ecological, interpersonal, conceptual, extended, and private aspects. I've suggested that the best way to think of these different aspects is not to consider them as aspects of a particular entity that we call 'self', but to think of them as forming a 'self-pattern' (Gallagher 2013). On this pattern theory of self, what we call 'self' is a pattern formed by a sufficient number of characteristic factors or processes (bodily processes, experiences, expressions, behaviors, actions, and so forth). To be clear, however, within the self-pattern there is no element that operates as an agent. As Scott Kelso (1997, 1) indicates, 'patterns in general emerge in a self-organized fashion, without any agent-like entity ordering the elements, telling them when and where to go'. In this regard, there is no self within a self-pattern; a self, of the sort that you are, and that I am, just is the pattern.

Many interesting and important questions arise in connection with the concept of self-pattern. In this chapter I'm going to skirt a number of major issues, such as what precisely a pattern is¹; what makes a particular pattern a *self*-pattern rather than something else; or precisely what elements one should include in a self-pattern. I'll stipulate on this latter question and simply provide a list of elements that can count as possible factors or processes that make up a self-pattern. This itself will be sufficient to motivate the issue that I want to address in this chapter, namely, concerns about the unity or coherence of the self-pattern.

Here, then is a tentative list of factors, processes, or characteristic features that could count as contributing to a self-pattern. The list is not necessarily

complete. The source of the list is a broad philosophical history that includes discussions of self and personal identity. For each element I offer some notes to indicate the scope of each process or factor (or set of factors).

- 1 **Bodily existence:** Core bio-systemic and autopoietic processes related to motoric, autonomic, endocrine, enteric, immune, interoceptive functions, and which also allow the overall system to maintain homeostasis necessary for survival, and to distinguish between itself and what is not itself – an extremely basic set of functions that both enable and constrain all kinds of animal behavior. Such processes involve or include the CNS, as well as sensory-motor processes that underpin the body-centered spatial frame of reference, which in turn grounds a first-person perspective and contributes to specifying possible actions in peripersonal space.
- 2 **Prereflective experiential aspects:** This includes prereflective self-awareness, a structural feature of first-person consciousness constrained by bodily factors; one's experiential life includes the sense of ownership (mineness) and the sense of agency, which can involve various sensory-motor modalities, such as proprioception, kinesthesia, touch, and vision. These aspects form the experiential core of what is sometimes called the minimal self (Gallagher 2000; Gallagher and Zahavi 2020, 2021; Zahavi 2007).
- 3 **Affective processes:** The fact that someone manifests a certain temperament or emotional disposition reflects a particular mix of affective factors that range from very basic and mostly covert or tacit bodily affects (e.g., hunger, fatigue, libido) to what may be a typical emotional pattern (Newen et al. 2015).
- 4 **Behavioral/action aspects:** Behaviors and actions make us who we are – behavioral habits and skills reflect, and perhaps actually constitute, our character. This is a classic view that goes back at least to Aristotle. John Dewey (1922, 24), for example, holds that habit forms our effective desires, furnishes us with working capacities, and rules our thoughts because 'it is so intimately a part of ourselves. It has a hold upon us because we are the habit'. This is a view that continues to be expressed in recent science: 'Habits may define who we are' (Verplanken and Sui 2019).
- 5 **Intersubjective factors:** Humans (possibly some non-human animals) are born with a capacity for attuning to intersubjective existence (Reddy 2008; Rochat 2011; Trevarthen 1979); at a certain point in social relations a more developed self-consciousness arises – a self-conscious recognition of oneself as being oneself as distinct from others, a sense of self-for-others (Mead 1913; Sartre 1956), and a sense of being part of a group or community. As Dennett expresses it: 'In fact, we wouldn't exist, as Selves "inhabiting complicated machinery" as Wegner vividly puts it, if it weren't for the evolution of social interactions requiring each human animal to create within itself a subsystem designed for interacting with others' (Dennett 2003, 47).

- 6 **Cognitive and psychological factors:** These are aspects emphasized in traditional theories of personal identity highlighting psychological continuity and memory (e.g., Shoemaker 2011), including one's conceptual understanding of oneself, beliefs, cognitive dispositions, as well as personality traits.
- 7 **Reflective capacities:** The ability to reflect on one's experiences and actions – closely related to notions of autonomy and moral personhood, including the capacity to reflectively evaluate and form second-order volitions about one's desires (Frankfurt 1982; Taylor 1989). Reflection does not necessarily mean introspection, typically understood as a kind of isolated, internal observation of mental states. Rather, reflection can involve self-interpretation; it can be directed at situations that involve worldly events (Evans 1982, 225); and it may even be conducted by engaging in conversation with others.
- 8 **Narrative aspects:** Self-interpretation has a narrative structure and recursively reflects (and often reinforces) the self-pattern. On some theories, selves are inherently or constitutively narrative entities (Schechtman 2011). The conception of this narrative aspect ranges from an abstraction (Dennett 1991b) to a complex accomplishment (Ricoeur 1992).
- 9 **Ecological relations**²: According to James (1890) we identify ourselves with our stuff – physical pieces of property, clothes, homes, and various things that we own, the technologies we use, the institutions we work in, etc. Our embodied-situated actions engage with (and sometimes incorporate) artifacts, instruments, bits, and structures of the environment in ways that define us and scaffold our identities. Situations shape who we are, and affordances define our possibilities. Autonomy in this affordance context can be viewed as a complex relational accomplishment.
- 10 **Normative factors:** Our extensive engagement with the environment also includes social and cultural practices. These are not just what we do, but involve what we ought to do, and obligations that we keep or not. Constraints (and sometimes well-defined roles) imposed by social, cultural, institutional factors shape our habitual behaviors, and our self-conceptions of who we are, and who we think we should be.

Such factors are variables that can take different values and weights in the dynamical constitution of a self-pattern. A pattern, however, exists not just as a collection of elements, but, importantly, involves a set of dynamical relations among these elements. The specific value and weight each element has in the pattern will depend on its relations with other elements. We can think of the pattern as a dynamical gestalt where if one factor (or value or weight relative to the whole) is changed above a certain threshold, some or all of the other factors (and perhaps the whole) adjust. In most, even if not all, cases this adjustment tends to maintain the whole as a continuing (albeit changing) pattern.

2 The issues: unity and coherence

A number of issues immediately arise on this view. One issue is this: if we say that different selves are constituted by different patterns, but a pattern can change over time – that it can develop or perhaps be transformed by certain events – then questions arise about unity, stability, and self-identity over time. Although a self-pattern is not a thing, entity, or substance, one still requires some kind of identity conditions. We need to account for the possibility of stability and continuity across changes in and of the pattern over time. A similar worry about unity or coherence is expressed, for example, by Miriam Kyselo (2014) who writes:

Once the diversity of self-related phenomena is acknowledged [as in the concept of self-pattern], we also need to understand how the elements of a collection of relevant self features interrelate. A pattern approach to the self acknowledges diversity but lacks integration, offering no account of the individual as explanatory whole.

(2014, 1; also see Beni 2016; de Haan et al. 2017)

As Kyselo rightly suggests, researchers in cognitive science as they set up experiments, and psychiatrists as they evaluate patients and attempt to understand pathologies of the self, need to have a conception of the self as an explanatory whole. On the one hand, it's not enough, of course, to simply say that the self-pattern is a dynamical gestalt (or what Kelso calls a 'dynamic pattern') and that the different elements of the self-pattern are dynamically related. On the other hand, this is not something that one can determine *a priori*. Rather, we need precisely the empirical and clinical studies mentioned by Kyselo to help specify how something like the dynamical integration of a self-pattern can occur.

The kind of pattern that we want to conceptualize is more than just an arrangement of elements. A self-pattern is not a static pattern like the wallpaper in your room, or a piece of abstract painting hanging in a museum. It's more like a piece of music being played by a quartet or orchestra. Perturbations or fluctuations are essential for the emergence of a specific order or pattern. Importantly, dynamical systems are unstable. We've known this about the brain, for example, for over a hundred years. An experiment by Layton and Sherrington (1917) demonstrated reversals and deviations of response in the body map of the motor cortex. Their point-by-point stimulation of this area showed fluctuating function – 'the point that had previously yielded elbow flexion as its primary movement, that point now yielded abduction of thumb as its primary movement ... and thumb and index movements as primary responses trespassed into the area that had previously yielded shoulder movement as the primary response' (1917, 141). For them, this demonstrated 'expression of the functional instability of a cortical motor point' (145). This doesn't mean that the brain is disorganized; it means it is dynamically organized and context sensitive (in the

sense that it is sensitive to repeated stimulation over time). 'It's not a static keyboard playing out a motor program' (Kelso 1997, 262) – it's more like a flexible stringed instrument where vibrations and bendings yield precise, or precisely imprecise patterns of sounds, or like vocalizations capable of nuanced intonations that produce improvised patterns of determinate meanings.

It may seem counterintuitive to philosophers who are used to working on questions that concern personal identity that instability may be more important than stability in what we are calling a self-pattern. Consider, for example, the angst that David Hume voiced when he was unable to discover anything stable in his experience because what we call self is supposed to be just that – something stable that remains identical over time.

If any impression gives rise to the idea of self, that impression must continue invariably the same, thro' the whole course of our lives; since self is suppos'd to exist after that manner.... [What we find, however, is] a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement.... [N]or is there any single power of the soul, which remains unalterably the same, perhaps for one moment. The mind is a kind of theatre, where several perceptions successively make their appearance; pass, re-pass, glide away, and mingle in an infinite variety of postures and situations. There is properly no *simplicity* in it at one time, nor *identity* in different; whatever natural propensity we may have to imagine that simplicity and identity. The comparison of the theatre must not mislead us. They are the successive perceptions only, that constitute the mind....

(1739, 251–253)

A self-pattern is neither a unitary or stable impression, nor a mere bundle of such impressions, since it includes more than mental impressions, and, in any case, is more than a bundle or aggregate; rather, it is characterized by an integration of heterogeneous factors. Some degree of instability in a pattern allows for change and flexibility such that the pattern rarely undergoes a radical break. Even though the self-pattern changes over time in most cases it remains a signature pattern. This means:

- 1 Despite changes in some features, the general parameters of a greater part of the self-pattern remain intact. For example, in changing social or normative contexts (among friends, among strangers, in the pub, at a religious service, at a football game, etc.) a person may engage in different behaviors of intersubjective interaction where different aspects of their personality and different affective profiles emerge. These changes may lead to small changes in some other aspects of the self-pattern, but the overall arrangement and the overall set of relations hold steady.
- 2 In the case of overall changes there is an observable continuity in the pattern. The best example of this is development across a life span.

There can be varying degrees of change in cognitive ability, personality, behavior, and even in bodily details, but what persists is some recognizable continuance in features or structure.

Judgments about such continuity through change are, of course, made by different stakeholders in different contexts and on different time scales. An encounter between old friends is one thing, raising a child, or living with a family member is something else, and our own self-perceptions are even more changeable within a very short timeframe, as Hume acknowledged (even to himself). There is also a certain recursivity built into such changes in the sense that such diachronic changes are part of what forms the self-pattern, and we would be surprised if no such changes occurred.

Just here philosophers want to raise questions about continuity, unity of self, and identity over time. If the self is a pattern, and the pattern changes, then doesn't the self change? Yes, as selves do. Paul Ricoeur (1992, 116) addresses this issue by making a distinction between *idem* identity (that which remains the same, persists across time) and *ipse* identity (that which makes me who I am in relation to another, even as I change over time). These are two different problematics addressing two different questions. As Marya Schechtman (1996) puts it, this is the distinction between the issue of reidentification – what makes a person the same from one time to another? – and the issue of characterization – what makes a person the person she is? To the extent that there are multiple questions about identity, how does the notion of a self-pattern, and specifically, certain features of the self-pattern, answer to the question about *ipse* identity or the issue of characterization?

In the following sections I argue that the narrative aspect of the self-pattern plays a special role in regard to this question about the unity or identity of the self-pattern. Self-narrative reflects or tracks all of the other aspects of the self-pattern. To get a fuller sense of what narrative encompasses, however, we can pursue the idea that the concept of self-narrative addresses what Ricoeur calls the issue of *ipse* identity, or what Schechtman calls the characterization question.

In the case of *idem* identity, which concerns reidentification or numerical identity, what makes the self one and the same thing, over time? Schechtman (1996, 67) makes a good case for bodily continuity as the criterion of reidentification, and although this continues to be controversial, I think it's the correct way to think about this question. In terms of a self-pattern, a unique living body (bodily existence) operates as the principle of *idem* individuation, a kind of ontological anchor that allows us to say this is the same person from one time or context to another, even when some other aspects of the self-pattern change. The idea is more nuanced than that since bodily existence also changes, although it does maintain some continuity over time; but for our purposes we need not enter any further into questions about *idem* identity.

Both Ricoeur and Schechtman point to narrative as a way to answer the characterization question of *ipse* identity. I think that this answer gets con-founded in reference to a third question which we can call the 'self-constitution

question'. Schechtman clearly distinguishes between the reidentification question and the characterization question, and these questions are answered, respectively, in terms of bodily continuity and narrative. In her discussion of narrative characterization, however, Schechtman drifts into a discussion of the following question: what is it that constitutes a person or self? The shift between these two questions takes place as one turns p. 93 of her book (1996). On this page she clearly takes narrative to be addressing the characterization question: what is it that makes a person who he is, even as he changes over time. This is a question about a person's identity, and Schechtman's view is that 'a person creates his identity by forming an autobiographical narrative – a story of his life' (93). She relates this directly to four concepts that are the concern of personal identity discussions: survival, moral responsibility, self-interested concern, and compensation. Once we turn to p. 94, however, she starts to answer a different question – the self-constitution question: what is it that makes some entity a person? Her answer, once more, is narrative. Indeed, she calls her position the 'narrative self-constitution view'.

At the core of [the self-constitution] view is the assertion that individuals constitute themselves as persons by coming to think of themselves as persisting subjects who have had experience in the past and will continue to have experience in the future, taking certain experiences as theirs.

(94)

Here, then, are the three questions that Schechtman distinguishes.

- 1 *Idem*/reidentification: what makes someone the same person over time?
- 2 *Ipse*/characterization: what is it that makes a person who she is, even as she changes over time?
- 3 Self-constitution: what makes an individual a person (rather than a non-person)?

Schechtman thinks that the body answers the first question, and narrative answers the second and third. Her answer to the third question, that narrative is definitive of personhood, however, leads to some contentious problems about claims such as 'to be a *person* at all one must construct a narrative [in the form of a conventional linear story]' (99), and to a later debate with Galen Strawson (2004) who argues against this sort of view. We can circumvent such contentious debates, however, simply by giving a different answer to question 3: namely, a person is constituted by a self-pattern. That answer does not require that every person has a strong self-narrative. Indeed, a particular person may lack narrative capacity, as sometimes happens in disorders involving dysnarrativa (Bruner 2002). We would not say, in such a condition, however, that this individual is not a person, since a significant part of the self-pattern still remains in place. What we can say in regard to the self-constitution question is that narrative may contribute to the constitution of the self-pattern, but is not a necessary condition.

The question I want to address in this chapter is the second question about *ipse* identity. Specifically, if narrative is the answer to this question, as both Ricoeur and Schechtman suggest, what does that tell us about narrative in the context of the self-pattern. Let me note that Ricoeur presents a nuanced view of the distinction between *ipse* and *idem* questions about identity insofar as they can be seen to overlap with respect to the notion of temporality. The concept of character, as it pertains to personal identity, is seemingly something that involves persistence over time, and a kind of ipseic consistency over time.³ Narrative, he suggests, operates as a kind of mediator between this overlapping aspect, and what he calls self-maintenance or self-constancy and the capacity to make a promise (1992, 118–119). In terms of the self-pattern, I think what’s true about this is that narrative reflects all the other aspects of the self-pattern (including bodily existence), not in a passive way, but in a way that in some sense mediates across their dynamical relations. We should not read this as suggesting that narrative solves the reidentification question. If we pursued that line of thought a third more radical view could be entertained, one would put Ricoeur’s position together with Schechtman’s and have narrative doing all the work in answer to all three questions. I don’t want to go that way. My focus will be to show how narrative works to answer the question about *ipse* identity by reflecting all other aspects of the self-pattern.

3 How narrative connects

We can understand the narrative aspect of self as a window that opens onto the self-pattern and offers a way to map the dynamical relations among the various factors in that self-pattern (Gallagher and Daly 2018). Self-narratives reflect, explicitly in content, and sometimes implicitly in form, all the other aspects of the self-pattern. In this section I want to trace the connections between narrative processes and other elements in the self-pattern. To do this I’ll bring together a number of considerations that get worked out in a variety of approaches to the concept of the narrative self. Even approaches that make narrative central to conceptions of self are unable to avoid pointing to the other factors beyond narrative that support the concept of a self-pattern.

As a first point of connection we should consider our *bodily existence*, our corporeal anchoring in the world. This is reflected in narrative insofar as narrative is primarily about *action*, and action is something accomplished bodily. Ricoeur, in his analysis of narrative and personal identity, makes this clear: ‘in virtue of the mediating function of the body as one’s own in the structure of being in the world, the feature of selfhood belonging to corporeality is extended to that of the world as it is inhabited corporeally’ (1992, 150). This linkage between body and world, between action and context, is reflected in self-narrative. Self-narrative tracks that linkage in a way that bestows on action the status of genuine action (i.e., as more than mere movement or behavior) – action, especially as it is tied to a practice. Action is never just bodily movement or ‘basic action’ (as defined in action theory [see Gallagher 2020]);

it's a complex embodied performance situated in circumstances that define its meaning. Precisely in this respect, the *intersubjective*, *ecological*, and *normative* features of the self-pattern are not isolated realms separated from embodied action. 'Practices are based on actions in which an agent takes into account, as a matter of principle, the actions of others' (Ricoeur 1992, 155). Actions are often interactions that involve other agents. Moreover, anything that might count as an individual practice is always derived from the practices of others.

In our experience the life history of each of us is caught up in the histories of others. Whole sections of my life are part of the life history of others – of my parents, my friends, my companions in work and in leisure. What we said above about practices and about the relations of apprenticeship, cooperation, and competition that they include confirms this entanglement of the history of each person in the histories of numerous others.

(Ibid, 161)

A practice, according to Ricoeur, may be a set of actions within a context that constitute a profession or a game – farming and chess playing are the examples he suggests. For example, 'shifting the position of a pawn on the chessboard is in itself simply a gesture, but taken in the context of the practice of the game of chess, this gesture has the meaning of a move in a chess game' (Ibid, 154). Something similar might be said of a chess piece, e.g., the rook. That is, what makes a rook are the rules of chess and the use to which it is put in the context of a game (see Haugeland 1998). But in the case of a human agent it is more than external *norms* and the *ecological context* of action that makes the action and the agent what they are. It is also what we might call the internal relations among the elements of the self-pattern that make the action *my* action, which may include the experiential sense of agency and reflection. This includes my ability to give a narrative account of that action.

Narrative competency is not simply the ability to comprehend and produce stories. It includes the reflective capacity to make reports on one's experiences and actions. It includes not just abilities for understanding narratives, but also capacities for narrative understandings, which allow us to frame our understanding of self and others in a narrative way, and thence to form/produce self-narratives and narratives about things, events, and other people. We can also distinguish between this narrative *framing*, as an implicit process or practice of seeing/understanding events in a narrative framework, and narrative *production*, which is the explicit construction of stories or narrative reports. When Jerome Bruner (1990) describes how we come to know our world and construct our representation of reality through the use of narrative, and when he suggests that 'we organize our experience and our memory of human happenings mainly in the form of narrative – stories, excuses, myths, reasons for doing or not doing, and so on' (4), he is describing narrative framing.

We make sense out of our own actions and out of the actions of others by placing them in a narrative framework. Narrative recounts actions, and their

reasons, as they figure against a larger history and set of projects. Ricoeur (1992) is close to the mark: I understand myself, and I encounter the other person, not abstracted from circumstances, but in the middle of something that has a beginning and that is going somewhere, and is part of a shared world. I see other people in the framework of a story about some action or practice in which either I have or do not have a part to play. I see their actions and my own actions, or possible actions, in what Bruner (1996) calls the ‘landscape of action’, which is constituted by embodied actions and the rich worldly contexts within which they act – contexts that operate as scaffolds for the meaning and significance of actions and expressive movements. The landscape of action relates directly to what Rietveld and Kiverstein (2014) call the ‘landscape of affordances’, which includes my *ecological possibilities* for interacting with others.

Our experience of others transforms our *self-experience* (the experience of our own possible actions), and narrative has a role to play in this. It is not just that, as Alistair McIntyre (1981, 212) puts it, ‘because we live out narratives in our lives and because we understand our own lives in terms of narratives that we live out, that the form of narrative is appropriate for understanding the actions of others’; rather, it begins by going the other way: because we first frame our understanding of the actions of others in narratives, the form of narrative is appropriate for the *reflective understanding* of ourselves.

This starts when we are young children. Since we develop in social contexts and normally acquire the capacity for narrative in those contexts, then the development of self-narrative obviously involves others. Katherine Nelson (2003) points out that ‘with respect to the child’s own experience, which is forecast and rehearsed with him or her by parents’, competency for self-narrative starts to emerge in two-year olds. Self-narrative requires building on our experiences of others and their narratives, so ‘children of 2–4 years often “appropriate” someone else’s story as their own’ (Nelson 2003). In addition, carving out one’s own character within a set of narratives requires a reflective awareness of having a point of view that is different from others. By the time infants are two or three years of age and well-practiced in understanding immediate environments and events as other people understand them, the acquisition of language, plus a relatively stable third-person, reflective grasp on themselves, as reflected in the capacity to recognize their own image in the mirror, feed a developing conceptual understanding of themselves that is essential to the onset of autobiographical memory.

By 18–24 months of age infants have a concept of themselves that is sufficiently viable to serve as a referent around which personally experienced events can be organized in memory.... The self at 18–24 months of age achieves whatever ‘critical mass’ is necessary to serve as an organizer and regulator of experience.... This achievement in self-awareness (recognition) is followed shortly by the onset of autobiographical memory

(Howe 2000, 91–92)

Autobiographical memory is one aspect that shapes narrative competency – the ability to see things in a narrative framework. Along with all the other developmental markers, autobiographical memory helps to kick-start narrative abilities during the second year of life. Two-year olds may start this process by working more from a set of short behavioral scripts than from full-fledged narratives; initially their autobiographical memories have to be elicited by questions and prompts (Howe 2000; Nelson 2003). From two to four years, children fine-tune their narrative abilities via further development of language ability, autobiographical memory, and a more stable self-concept.

Narratives – specifically narratives about others that start to shape our own self-narratives – help to establish and reinforce the *normative factors* that become part of the self-pattern. When children listen to stories (and when, as adults, we are exposed to parables, plays, myths, novels, our favorite rap music, films, and other media) they become familiar with characters in a range of ordinary or extraordinary situations, and the sorts of actions appropriate to such characters, all of which helps to shape their expectations about others and about themselves (Gallagher and Hutto 2008; Nelson 2003; Richner and Nicolopoulou 2001). Children often incorporate such norms in pretend play or play acting.

Narratives thus contribute to our normative understanding of what others can expect from us, and what we can expect from others in certain situations. They provide an important constraint on what actions are acceptable and what are not acceptable. Narratives can form part of the structure that shapes social and cultural institutions (Gallagher and Tollefsen 2019), which in turn inform our *intersubjective interactions*. Through them we learn the norms associated with social roles that pervade our everyday world and our social practices, in homes, schools, playgrounds, shops, restaurants, etc. These are the normative aspects that shape our self-identity.

Developmentally, in our narrative understanding of others, we begin to shape our own self-narrative, registering not only their actions and attitudes but also our own experiences and, at the same time, in a way that differentiates self and other. We become, in Charles Taylor's (1985) phrase, 'self-interpreting animals' because we are intertwined with others. Other larger, cultural narratives that help to constitute a community's shared normative practices and our common-sense understandings also shape our self-understandings and our understanding of others.

To summarize:

- 1 Narrative is primarily about actions, which are often intersubjective interactions and often involve social-cultural practices;
- 2 Actions and intersubjective interactions are embodied;
- 3 Because our bodies are always in-the-world, our actions are always ecologically situated;
- 4 Action-situations involve extended, worldly aspects, but especially other people and social practices constrained in a normative fashion;

- 5 Narrative reflects such structures and allows us to frame our understanding of both our own actions and those of others in ways that recursively shape those actions and our own experiences;
- 6 The development of narrative competency and self-narrative begins at a young age reflecting and contributing to the development of our linguistic and cognitive abilities, fostering the development of episodic-autobiographical memory and self-concept.

4 Narrative identity: fiction or reality

Someone might object: this is all well and good, but this kind of narrative identity, this making of coherency, or some degree of coherency, in the self-pattern is still only a kind of construction; it has no real depth and may be a kind of fiction, as suggested by thinkers like Hume (1739) or Dennett (1991b). One response to this concerns the role of affect. In what I've said so far, narrative has hit on every aspect of the self-pattern – bodily, experiential, behavioral/action-related, intersubjective, psychological-cognitive, reflective, ecological, and normative – except affectivity. Schechtman (1996, 97) makes it clear, however, that affect is an important part of this mix of elements: 'We expect a person's beliefs, desires, values, emotions, actions and experiences to hang together in a way that makes what she says, does, and feels psychologically intelligible'. Schechtman introduces the idea that affect is something like the glue that keeps these various elements hanging together, and she does this through her interpretation of John Locke's classic account of personal identity. In contrast to interpretations that emphasize the role of memory, i.e., psychological continuity, as the basis for personal identity, Schechtman emphasizes that Locke

stresses the *affective* side of consciousness. He paints a picture of consciousness as the faculty whereby we experience pleasure and pain, happiness and misery. He tells us, for instance, that '*Self* is that conscious thinking thing ... which is sensible, or conscious of Pleasure and Pain, capable of Happiness or Misery, and so is concern'd for it *self* as far as that consciousness extends', thus emphasizing the definition of identity in terms of sameness of consciousness precisely because it is in consciousness that we experience the affect that underlies self-interested concern, compensation, and justice of punishment.

(Schechtman 1996, 108–109)

Affect, in this sense, is what makes one's body one's own, providing for a sense of body-ownership. On the same basis, we own our past and present actions. Affect, in this regard, operates as a glue that binds narrative identity to bodily existence and action. Schechtman makes clear that affect does not always work in a conscious way. The affective associations connected with past actions condition our current actions, sometimes even if we do not have

explicit memories of those past actions. More generally, Schechtman points out, the affective dimensions that are part of the way we have been raised may contribute to our feelings of worthiness and self-esteem. 'It is no great revelation that a person who feels loved and valued by a stable family in childhood is more likely to grow up feeling worthy, entitled, and secure than a person who is made to feel worthless and incompetent' (111). Importantly, this general orientation toward the world will affect a person's interactions with others, her choice of action, and her specific dispositions in various situations. All of this, in turn, is reflected in one's narrative, since these affective experiences

give us a "script" – a sense of self, an idea of who we are and what kind of story we are living.... This, then, is how Locke's insight can be used to yield a helpful understanding of what is involved in having a narrative self-conception.... To have a narrative self-conceptionis thus to experience the events in one's life as interpreted through one's sense of one's own life story, and to feel the affect that follows upon doing so.

(Schechtman 1996, 111–112)

I interpret Ricoeur to be offering a similar conception of the relation between self-narrative and affect (a relation that makes self-narrative personal). Ricoeur makes it clear that there is an unusual causal relation between narrative and the events narrated, which involves 'action dynamics' (Ricoeur 1992, 144). It's possible, he suggests, to give a causal, impersonal account of a series of events or actions. A narrative, on its own, does not necessarily add extra causal forces to the series of events. At the same time, however, a narrative can transform the meaning of an event; it can have an affective effect. Narrative provides extended context, and for that reason, the event narrated can take on a certain personal (and this means 'affective') significance that it otherwise would not have. More specifically,

by entering into the movement of a narrative which relates a character to a plot, the event loses its impersonal neutrality. By the same token, the narrative status conferred upon the event averts the drift of the notion of event which would make it difficult, if not impossible, to take the agent into account in the description of the action.

(142n1)

Ricoeur uses the example of the promise, but we can see the same dynamics, and the importance of affect, at work in the example of regret. Regret involves a form of narrativ self-knowledge in which I, as the narrator, can give an account and evaluate my past action. Sometime in the past I made the wrong choice. Now I come to regret that choice, and this manifests itself in my self-narrative. Specifically, what gets reflected in my narrative are changes or adjustments in my self-pattern. Frith and Metzinger (2016) show

how this works in regard to the minimal experiential aspects of the self-pattern, which include the sense of agency and the sense of ownership. They point to a changing status in the sense of agency as I relate, by way of regret, to my past action. When I acted, I had a sense of control over my action since I made a choice to act in that way. This sense of agency does not apply to my regret, however; I seem unable to avoid my regret about that action. Accordingly, my narrative reflects adjustments (a comparative decrease) in my sense of agency. In contrast, I have a continued sense of ownership for both my action and my regret. Thus:

while the sense of agency is represented as something we possessed in the past, the state of regret itself does not itself involve a sense of agency. While the phenomenology of ownership is crisp and distinct (I identify with my regret, it is an integral part of myself), regret itself is not an action. It is a kind of inner pain that simply appears in us.

(Frith and Metzinger 2016, 202)

The emergence of the inner pain (a bodily and emotional phenomenon) reflects an affective binding in the self-pattern – ‘not something we can distance ourselves from – another important way in which regret involves a loss of [agency]’ (Ibid).

My regret is not just a characteristic that I can narrate, it is something that takes over my narration.

The phenomenology of regret can be described as a loss of control over our personal narrative, and in this sense it is also a threat to our integrity [and sense of agency]. It is a threat to the integrity of our autobiographical self-model, because, on the personal level of description, we become aware of an irrevocable damage to our life narrative.

(Ibid)

The relation between my action and my regret, despite variations in my sense of agency, means that I am the same person who acted in the past, and continues to be responsible for my actions. I think that Frith and Metzinger are wrong to suggest that this is a fictional unity: ‘a transtemporal, fictional “self” constituted by the narrative (Frith and Metzinger 2016, 204n2). Rather, one should say that this is a very real unity or coherence because the action and the affective states – the pain and the regret – are real, and are factors that modulate my self-narrative and specify my self-pattern.

It is also clear how intersubjective and social dimensions are integrated in these concerns about both agency and affect, since others may be affected (for example, frustrated) by my actions, which is also something the narrative may reveal. The narrative may in fact reflect my identification with the values and desires of others, which may be what leads to the regret and which may increase my emotional suffering (see Frith and Metzinger 2016,

204–205). Accordingly, we can say that in the dynamical adjustments in the self-pattern intersubjective/social and normative factors determine my reflective evaluation and my feelings about what I have done. Expressing this in my self-narrative may influence my present and future behavior, or have an effect on my future choice of action, which again will be reflected in my ongoing self-narrative. Our actions often come to be guided by our self-narratives, which prospectively can enter into intention formation through reflective practices (Velleman 2006). In such a case there are clear links made to ‘cultural practices, such as moral codes and laws. By generating beliefs about self-control [my narrative] shapes the sense of self and gives rise to concepts like responsibility, intentionality, accountability, culpability, and mitigating circumstances’ (Frith and Metzinger 2016, 205). My action, my regret, my pain, mediated in my narrative practice, are here making a loop through normative factors that ultimately feeds back to influence my behavior.⁴

This kind of affective binding, mediated by self-narrative, reflects the dynamical relations that define the double role of narrative in the self-pattern. Self-narrative not only is reflective of the other processes of the self-pattern, but it is itself one component of the self-pattern. In other words, narrative is not a one-way, disinterested reflective formatting of the self-pattern (as if operating as an external observer), it is affected by and shaped by all those other factors, and is affected when those other factors become discordant.

One can see this clearly in the case of depression. Depression involves bodily changes (fatigue, insomnia, increased interoceptive awareness), increased self-focus, and negative emotions, such as worry, guilt, shame, jealousy, and increased anxiety (Fingelkurts and Fingelkurts 2017, 30). Such changes also involve intersubjective relations; a person with depression may feel like a burden for others, excluded, and even a profound intersubjective alienation. Cognitive changes may include disordered attention, inability to concentrate, indecisiveness, recurrent thoughts of death/suicide, excessive rumination, toxic thought processes, mind fog, changed time perception, and an inability to imagine a different future. Likewise, there are changes in ecological affordances; things and surroundings are experienced as less salient; diminished engagement with the world; loss of sense of belonging or fitting in place. Some of these real changes in the self-pattern may be explicitly represented in the person’s narrative – that is, the person may report that they are suffering insomnia, feeling guilty, feeling like a burden, and so forth. But also, such changes have a real effect on the structure of the self-narrative. Patients with major depressive disorder may repeatedly rescript conversations that were deemed unsatisfactory; there is a predominance of the use of first-person pronouns in their narratives; not only are their narratives about the past couched in terms of loss, failure, and damage, but present narratives hold little or no interest, and future narratives have dried up.

5 Conclusion

The self-pattern is heterogeneous. It emerges from heterogeneity as it dynamically integrates very diverse phenomena. The integration brings some kind of coherence that is reflected in self-narrative as a form of *ipse* identity. This does not mean that there is no discordance in self-narrative, since a self-pattern can be less or more discordant, as one sees in the extremes of psychopathology. Likewise, the ‘narrative unity of life’ (MacIntyre) is not the totality of a life. Our narrative is always incomplete, and not always as unified as we might think. Ricoeur calls it the ‘narrative incompleteness of life’ due, in part, to our entanglement with others’ life histories (1996, 161), and to the extent that one is never the sole narrator of one’s story.

As Katrina MacKenzie rightly suggests, when narratives explain the connections between the actions they recount, ‘they give shape and coherence to our lives, or at least to the various sequences that make up our lives...’ (2007, 268–269), and they inform future actions, providing motivational, normative, and intentional connections. In addition, more implicitly, narratives, to the extent that they reflect norms and customs as well as personal experiences, may start to inform our behaviors without our being aware of it. My actions, my affordances, what I can do, what my possibilities are, are shaped by the effects of my own narrative and my normative understanding of what is possible.

It is in this sense that narrative not only reflects, but contributes to, by being part of, the self-pattern. My self-narrative can add real coherence to the self-pattern to the extent that it pulls together my actions, my sense of agency, my reflective abilities, the affordances offered by the environment, the normative forces that constrain my life, and my intersubjective relations. Narrative, accordingly, has a significant role to play in providing an account of *ipse* identity, answering the characterization question, without, however, being the answer to the constitution question. The answer to that question is the self-pattern itself.

Notes

- 1 The nature of a pattern is a topic I won’t discuss here (see Dennett 1991a; Hauge-land 1998; Kelso 1995). The idea has some affinity to Morton Beckner’s (1959, 22) definition of a polytypic concept. Thanks to Kenneth Schaffner for pointing this out. Also see Schaffner (1993, Ch. 3). One important difference, however, is that Beckner defines the polytypic concept as an aggregate; in contrast, a pattern, as understood here, is integrated by dynamical relations and is therefore something more than an aggregation of items. I characterize it as a dynamical gestalt. The idea is close to the concept of *dispositif* (Foucault 1980) or *agencements* (Deleuze 2006, 179; Deleuze and Guattari 1987; see Slaby et al. 2019), a heterogeneous ensemble of components that co-function as a coherent unity defined in terms of the relations between its integrated elements.
- 2 The aspect I am calling ‘ecological’ here is not the same as Neisser’s concept of ecological, which he draws from Gibson’s ecological psychology. Neisser means the ecological self-awareness involved in proprioception/kinesthesia that gives us

a direct embodied sense of self-movement distinct from other movement in the environment. On the concept of self-pattern outlined here, Neisser's notion of ecological would be included in the experiential aspect. The concept of 'ecological relations' mentioned here does include the idea of affordance which is also drawn from Gibson (1977).

- 3 Ricoeur describes it as partaking 'simultaneously of two orders, that of objectivity and that of existence' (1992, 119n4). Ricoeur, however, in the context of his discussion of personal identity moves away from this overly uniform and permanent characterization of character, turning toward a more Aristotelian notion of an acquired disposition or habit that is not unrelated to social and normative factors, which in turn may 'place a "cause" above our own survival' (120).
- 4 The looping is complex. Somogy Varga (2015, 71) writes:

not only does autobiographical self-reflection involve the co-construction of what is recollected, establishing a link between the [narrator's] current view of himself and a set of past experiences, but also this process occurs in a loop-like, dynamic way – what is recollected influences the present self-understanding, yet the recollected material is itself influenced and altered on the basis of the present self-understanding.

References

- Beckner, M. (1959). *The Biological Way of Thought*. Berkeley: University of California Press.
- Beni, M.D. (2016). Structural realist account of the self. *Synthese* 193: 3727–3740.
- Bruner, J. (2002). *Making Stories: Law, Literature, Life*. Cambridge, MA: Harvard University Press.
- Bruner, J.S. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.
- Bruner, J.S. (1996). *The Culture of Education*. Cambridge, MA: Harvard University Press.
- de Haan, S., Rietveld, E., Stokhof, M. and Denys, D. (2017). Becoming more one-self? Changes in personality following DBS treatment for psychiatric disorders: experiences of OCD patients and general considerations. *PloS One* 12 (4): e0175748.
- Deleuze, G. (2006). *Two Regimes of Madness*. ed. D. Lapoujade. Cambridge: MIT Press.
- Deleuze, G. and Guattari, F. (1987). *A Thousand Plateaus Capitalism and Schizophrenia* (Trans. B. Massumi). Minneapolis and London: University of Minnesota Press.
- Dennett, D.C. (1991a). Real patterns. *The Journal of Philosophy* 88 (1): 27–51.
- Dennett, D. (1991b). *Consciousness Explained*. London: Allen Lane.
- Dennett, D.C. (2003). Review of *Making ourselves at home in our machines: the illusion of conscious will* by Daniel Wegner. MIT Press 2002. *Journal of Mathematical Psychology* 47(1): 101–104.
- Dewey, J. (1922). *Human Nature and Conduct: An Introduction to Social Psychology*. New York: Henry Holt and Company.
- Evans, G. (1982). *The Varieties of Reference*. Oxford: Clarendon Press.
- Fingelkurts, A.A. and Fingelkurts, A.A. (2017). Three-dimensional components of selfhood in treatment-naïve patients with major depressive disorder: a resting-state qEEG imaging study. *Neuropsychologia* 99: 30–36.
- Foucault, M. (1980). The confession of the flesh. *Power/Knowledge. Selected Interviews and Other Writings, 1972–1977*. ed. C. Gordon. New York: Pantheon Books, 194–228.

- Frankfurt, H. (1982). The importance of what we care about. *Synthese* 53 (2): 257–272.
- Frith, C.D. and Metzinger, T. (2016). What's the use of consciousness. *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science*. eds. A.K. Engel, K.J. Friston and D. Kragic. Strüngmann Forum Reports, vol. 18. Cambridge: MIT Press, 193–214.
- Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences* 4 (1): 14–21.
- Gallagher, S. (2013). A pattern theory of self. *Frontiers in Human Neuroscience* 7 (443): 1–7. doi: 10.3389/fnhum.2013.00443.
- Gallagher, S. (2020). *Action and Interaction*. Oxford: Oxford University Press.
- Gallagher, S. and Daly, A. (2018). Dynamical relations in the self-pattern. *Frontiers in Psychology* 9: 664. doi: 10.3389/fpsyg.2018.00664.
- Gallagher, S. and Hutto, D. (2008). Understanding others through primary interaction and narrative practice. *The Shared Mind: Perspectives on Intersubjectivity*. eds. J. Zlatev, T. Racine, C. Sinha and E. Itkonen. Amsterdam: John Benjamins, 17–38.
- Gallagher, S. and Tollefsen, D. (2019). Advancing the 'we' through narrative. *Topoi* 38 (1): 211–219. doi: 10.1007/s11245-017-9452-1.
- Gallagher, S. and Zahavi, D. (2020). Phenomenological approaches to self-consciousness. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/>.
- Gallagher, S. and Zahavi, D. (2021). *The Phenomenological Mind* (3rd edition). London: Routledge.
- Gibson, J.J. (1977). The theory of affordances. *Perceiving, Acting, and Knowing*. eds. R. Shaw and J. Bransford. Hillsdale, NJ: Erlbaum, 67–82.
- Haugeland, J. (1998). *Having Thought: Essays in the Metaphysics of Mind*. Cambridge, MA: Harvard University Press.
- Howe, M.L. (2000). *The Fate of Early Memories: Developmental Science and the Retention of Childhood Experiences*. Washington, DC: American Psychological Association.
- Hume, D. (1739). *Treatise of Human Nature*. Oxford: Oxford University Press, 1978.
- James, W. (1890). *The Principles of Psychology*. New York: Basic Books, 1950.
- Kelso, J.S. (1997). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT press.
- Kyselo, M. (2014). The body social: an enactive approach to the self. *Frontiers in Psychology* 5. doi: 10.3389/fpsyg.2014.00986.
- Layton, A.S.F. and Sherrington, C.S. (1917). Observations on the excitable cortex of the chimpanzee, orangutan and gorilla. *Quarterly Journal of Experimental Physiology* 11: 135–222.
- Mackenzie, C. (2007). Bare personhood? Velleman on selfhood. *Philosophical Explorations* 10 (3): 263–281.
- McIntyre, A. (1981). *After Virtue: A Study in Moral Theory*. Notre Dame: University of Notre Dame Press.
- Mead, G.H. (1913). The social self. *Journal of Philosophy Psychology and Scientific Methods* 10: 374–380.
- Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology* 1: 35–59.
- Neisser, U. (1991). Two perceptually given aspects of the self and their development. *Developmental Review* 11(3): 197–209.
- Nelson, K. (2003). Narrative and the emergence of a consciousness of self. *Narrative and Consciousness*. eds. G.D. Fireman, T.E.J. McVay and O. Flanagan. Oxford: Oxford University Press, 17–36.

- Newen, A., Welpinghus, A. and Juckel, G. (2015). Emotion recognition as pattern recognition: the relevance of perception. *Mind & Language* 30 (2): 187–208.
- Reddy, V. (2008). *How Infants Know Minds*. Cambridge, MA: Harvard University Press.
- Richner, E.S. and Nicolopoulou, A. (2001). The narrative construction of differing conceptions of the person in the development of young children's social understanding. *Early Education and Development* 12: 393–432.
- Ricoeur, P. (1992). *Oneself as Another* (Trans. K. Blamey). Chicago, IL: University of Chicago.
- Rietveld, E. and Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology* 26 (4): 325–352.
- Rochat, P. (2011). What is it like to be a newborn? *The Oxford Handbook of the Self*. ed. S. Gallagher. Oxford: Oxford University Press, 57–79.
- Sartre, J.-P. (1956). *Being and Nothingness: And Essay on Phenomenological Ontology*. Paris: Gallimard, 1943; Transl., New York: Washington Square Press.
- Schaffner, K.F. (1993). *Discovery and Explanation in Biology and Medicine*. Chicago, IL: University of Chicago Press.
- Schechtman, M. (1996). *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
- Schechtman, M. (2011). The narrative self. *The Oxford Handbook of the Self*. ed. S. Gallagher. Oxford: Oxford University Press, 394–416.
- Shoemaker, S. (2011). On what we are. *The Oxford Handbook of the Self*. ed. S. Gallagher. Oxford: Oxford University Press, 352–371.
- Slaby, J., Mühlhoff, R. and Wüschner, P. (2019). Affective arrangements. *Emotion Review* 11: 2–12. doi: 10.1177/1754073917722214.
- Strawson, G. (2004). Against narrativity. *Ratio* 17(4): 428–542.
- Taylor, C. (1985). Self-interpreting animals. *Philosophical Papers: Volume 1*. Cambridge: Cambridge University Press.
- Taylor, C. (1989). *Sources of the Self: The Making of the Modern Identity*. Cambridge, MA: Harvard University Press.
- Trevarthen, C. (1979). Communication and cooperation in early infancy: a description of primary intersubjectivity. *Before Speech*. ed. M. Bullowa. Cambridge: Cambridge University Press, 321–347.
- Varga, S. (2015). *Naturalism, Interpretation, and Mental Disorder*. Oxford: Oxford University Press.
- Velleman, J.D. (2006). *Self to Self: Selected Essays*. New York: Cambridge University Press.
- Verplanken, B. and Sui, J. (2019). Habit and identity: behavioral, cognitive, affective, and motivational facets of an integrated self. *Frontiers in Psychology* 10: 1504.
- Zahavi, D. (2007). Self and other: the limits of narrative understanding. *Royal Institute of Philosophy Supplement* 60: 179.

Part 3

Intersubjectivity

9 Animals in person space

Marya Schechtman

In Nathaniel Hawthorne's short story *Wakefield*, the narrator recounts a story recollected from "some old magazine or newspaper" about a man he calls Wakefield, who left his home one day and did not return for over twenty years. As it turns out, for some perverse reason Wakefield had been living all that time right around the corner from the home he left, observing his abandoned household. After more than two decades, we are told, "his estate settled, his name dismissed from memory, and his wife, long, long ago, resigned to her autumnal widowhood – he entered the door one evening, quietly, as from a day's absence..." (Hawthorne, 1982, p. 290). The narrator concludes that Wakefield cannot resume his earlier life, or indeed any life at all, saying: "Stay, Wakefield! Would you go to the sole home that is left you? Then step into your grave!". He explains that

Amid the seeming confusion of our mysterious world, individuals are so nicely adjusted to a system, and systems to one another, and to a whole, that by stepping aside for a moment, a man exposes himself to a fearful risk of losing his place forever.

(Hawthorne, 1982, p. 298)

Wakefield is presented in the end as a ghost who can do no more than haunt the life he once lived.

Wakefield offers a powerful expression of the idea that the space we occupy in systems of social interaction – systems that engage with one another and with systems of nature – is central not only to our pleasure and well-being, but to our very integrity and persistence as individuals. This is something I have tried to argue in a philosophical context in a view I call the "Person Life View" (PLV),¹ which is offered as an alternative to both traditional psychological accounts of identity and biological or "Animalist" views. According to PLV, occupying and maintaining a place in a network of social interactions – a "person space" – is fundamental to personhood and personal persistence. This chapter seeks to clarify the notion of "person space" and its role in the identity and persistence conditions of beings like us, human persons, by addressing an apparent dilemma for the view. Since it claims

that social factors play a constitutive role in the individuation and persistence of human persons, PLV seems vulnerable to charges that it is objectionably conventionalist. If it avoids this worry by defusing or constraining the role of these social features, it begins to look as if it is just a disguised or complexified version of Animalism after all. Showing how the view can steer between the horns of this dilemma will help clarify the nature and role of person space, and reveal fundamental differences between the metaphysical starting point of this view and that shared by Animalism and traditional psychological accounts of personal identity.

I begin with a quick review of the debate between psychological and biological theorists into which PLV is an entry. Next, I give a brief overview of the account, as well as the worries about conventionalism it is likely to engender. Describing the strategy for avoiding these worries leads to the second horn of the dilemma, the possibility that the view is a disguised form of Animalism. I respond to this worry by showing how PLV rests on a different metaphysical perspective than that assumed by Animalism, and more broadly in the debate between Animalists and traditional psychological theorists.

1 The debate between biological and psychological accounts

For much of the 20th century, the predominant view of personal identity in analytic philosophy was the psychological approach. Rooted in John Locke's account (Locke, 1975, pp. 328–48), present-day psychological theorists (see, e.g., Lewis, 1983; Noonan, 2003; Parfit, 1984; Perry, 1972; Shoemaker, 1984) argue that personal identity is constituted by psychological relations – what makes a person at time t^2 the same person as a person at time t^1 is that the person at t^2 is psychologically connected to the person at t^1 in the right kind of way. The psychological approach is contrasted with the biological approach, which says that what constitutes the numerical identity of beings like us over time is sameness of human animal rather than psychological continuity. In rejecting the biological approach, which is in some respects more common-sensical, psychological theorists thus follow Locke in distinguishing between persons and human animals (for Locke, “sameness of person” and “sameness of man”). A person, for Locke, is “a thinking intelligent Being, that has reason and reflection, and can consider itself as itself, the same thinking thing in different times and places...”. He says also that “person” is “a Forensick term” (Locke, 1975, p. 335). In other words, a person is a psychological entity that possesses certain kinds of practical capacities and serves as a unit for the purposes of particular kinds of practical interactions and judgments, e.g., assessments of moral responsibility or of prudential reasoning.

The argument for the psychological approach depends upon hypothetical cases in which psychological continuity comes apart in imagination from continuity of human animal. It is argued that such cases support the idea that when psychological and biological continuity diverge, the *person* goes with

the former rather than the latter. Locke famously offers a case where the consciousness of a prince enters the body of a cobbler and argues that “everyone sees” that the resulting person “would be the same Person with the Prince, accountable only for the Prince’s Actions” (Locke, 1975, p. 340). Present-day theorists employ similar kinds of cases with a more science-fictional twist – imagining, for instance, brain transplants or brain state transfer devices, or replicators, or teleporters. The fundamental structure of the argument is essentially the same. We imagine psychological continuity coming apart from continuity of human animal and consider which kind of continuity person-related judgments like those of moral responsibility for past actions or rational self-interest for future states follow. The strong intuition on which psychological approaches rely is that in such cases the practical judgments follow psychological rather than biological continuation. This leads them to conclude that personal identity consists in psychological rather than biological continuity.

Despite the popularity and intuitive appeal of the psychological approach, it is not without detractors and there have been many defenders of the biological approach (see, e.g., Carter, 1989; Inwagen, 1990; Mackie, 1999; Snowdon, 2014). One influential example of such a defender is Eric Olson (e.g., Olson, 1997). In developing his version of this account, Olson offers a detailed set of arguments against the psychological approach which is especially helpful for our purposes. The central thrust of these arguments is that if we take seriously the idea that a person is a fundamental kind of entity, distinct from a human animal and possessing its own persistence conditions, we will be faced with extremely counterintuitive implications. These can be seen in his “fetus problem”, “vegetable problem”, and “too many thinkers” problem.

All three problems employ the assumption that a human animal is an entity with biological persistence conditions. If a person is a distinct entity with psychological persistence conditions and we are fundamentally persons as the psychological approach claims, none of us could ever have been a human fetus, since a fetus is not yet a person as psychological theorists define the term. Moreover, when a human animal has developed to the point where the features that make a Lockean person are expressed, psychological theorists are forced to say that a new, second entity, materially coincident with the human animal that has been there for a while, suddenly springs into existence. The view also has the implication that none of us could fall into a vegetative state, since a human in a vegetative state is not a Lockean person. Again, since the human animal clearly continues in the vegetative state, this means that when a human person falls into such a state an entity (the person) ceases to exist without any obvious material change. Finally, during the period where there is a human person, the psychological theorist will have to allow that the two beings – the person and the human animal – are materially coincident. This suggests that every thought (that involves brain activity) or utterance (that involves vocalization) must be attributed to both the human and the person, meaning each thought has two thinkers, and each utterance two speakers.

When we focus on these implications, Olson argues, the psychological approach seems needlessly complicated. Rather than thinking of the person as a separate entity, it makes more sense to think of personhood as an attribute that humans typically possess for much of their lives. To be a person, according to the psychological approach, is to have a particular set of capacities – for reason, reflection, agency, and so on. Why not, then, simply say that there is one entity, a human animal, that becomes a person when it gains these attributes and can cease to be a person if they are lost, but which typically exists before becoming a person and can survive the loss of personhood? This immediately avoids the awkward implications of viewing the person as a metaphysically distinct entity. Olson is happy to concede that the attribute of personhood is extremely important to most of us, so important that many do not care whether they persist at all if they do not persist as persons. What he wants to insist upon, however, is that we *can* persist as non-persons whether we find this interesting or valuable or not.

Olson also recognizes the strength of the intuition elicited by hypothetical cases like Locke's prince and cobbler or the present-day brain transplant cases. It is this intuition that tempts us to think that a person is an independent entity with its own persistence conditions. But we need not interpret these cases as psychological theorists interpret them, Olson argues. We can acknowledge and validate the central intuition uncovered in these cases without accepting a psychological account of identity. What these cases really show, he argues, is that when, e.g., the psychological life of the prince is transferred to the animal who is the cobbler, replacing the cobbler's own psychology, the resulting person is responsible for the prince's activities, and is the person the princess and queen should treat as they treated the prince. The intuition, then, is that the practical relations and judgments that usually attach to sameness of person would follow psychological rather than biological continuity should they come apart. This is not yet, however, to show that *we* go where the psychological continuity goes. This conclusion follows only if we assume that the person-related judgments and relations that usually track our identity *must* do so.

Olson suggests that we consider instead that the tight connection that usually holds between personal identity and practical judgments might cease to do so in the hypothetical cases described. If the psychological life of the prince enters the body of the cobbler, according to Animalism, the resulting person is still the cobbler – numerically the same entity that was unquestionably the cobbler before the psychological transformation. What the case reveals, however, is that in such odd cases practical judgments that typically track facts about identity no longer do so. It is in the light of these kinds of considerations that Olson claims that "it is fair to say that no account of our identity has yet been proposed that guarantees ... the coincidence of what is important in our identity with the actual conditions of our identity" (Olson, 1999, p. 165).

The dispute between psychological and biological accounts of identity is ongoing, and both views continue to develop in a wide variety of directions.

I think, however, that there is a radically different way to think about these matters that can speak to the powerful objections Olson raises against the psychological continuity theory but does, in its own way, guarantee coincidence between what is important in identity and its actual conditions.

2 The Person Life View

The PLV takes a page from Animalism by defining our identity in terms of the unity of a life. Insofar as Animalism sees us as fundamentally human animals, it defines our persistence in terms of the continuity of a biological life; one of us starts when the biological life of a human animal begins and ends when it ends. PLV makes a similar claim, except that the life it describes is not just a biological life, but rather what I call a “person life” which is a characteristic kind of life lived by persons. One of the key features of a person life is that it is lived within, and structured by, social and cultural institutions and practices concerning interactions among its participants. This notion obviously needs more unpacking than I can give it here, but it should be possible to convey enough of the central idea for present purposes. In the interest of space, I will focus here on PLV as an account, specifically, of the nature and persistence conditions of human persons.²

A quick way to get a sense of the strategy behind the view is to show how it is continuous with, and deviates from, traditional psychological approaches. Such psychological approaches, we have seen, start from Locke’s picture of “person” as a forensic term and emphasize the capacity for complex forms of interpersonal interaction and agency. PLV suggests that if we take this idea seriously, we will have to include social features in our account. There can be no interpersonal interactions without other people, and there cannot be the kinds of complex interactions and relationships on which psychological theorists focus without sociocultural institutions to support them. One of the key differences between the ways in which humans (who are the paradigmatic persons in most philosophical discussions) and other animals (those seen as non-persons in this schema) live is the way in which their interactions are codified in and conditioned by such institutions.

Locke and those who follow him define personhood in terms of the cognitive capacities that allow for these person-specific interactions, but PLV takes this basic notion of personhood in a somewhat different direction. It grants that persons must typically possess the kinds of cognitive capacities described in the Lockean view if they are to develop and maintain social and cultural institutions but sees personhood as consisting in living within these institutions, rather than possessing the capacities that allow for their creation. This involves two important shifts from standard psychological approaches. First is the already-mentioned emphasis on the life lived rather than on the capacities possessed, and so from individual attributes to social situation. This leads to the second shift, which is a more inclusive notion of personhood. On the standard view, an individual becomes a person only when she achieves

the higher-level cognitive capacities needed for moral agency and institution building (infants, for instance, are not yet persons). Someone can cease to be a person on the standard approach if she loses these capacities (those with severe dementia, for instance, are no longer persons), and those who never develop these capacities cannot be persons at all.

Since PLV focuses on the whole of a life lived within sociocultural institutions, it includes as person defining many kinds of interactions that are part of such institutions beyond the attribution of moral responsibility or assessments of prudential reasoning. There are myriad characteristic interactions between people over the course of their lives that require social institutions, and so require that persons typically possess higher-order cognitive capacities, but which do not require the possession of these capacities by everyone involved in them. We interact with infants as persons, for instance, when we dress them up and hold ceremonies to induct them into a religious tradition, count them for the census, register their birth, play peekaboo with them, or sing to them before dressing them for bed and putting them in a crib with sheets. We interact with dementia patients as persons when we visit them at the long-term care facility, take them to music therapy, or celebrate their birthdays. We interact with other people in this way when we put on team colors and watch a football match together, or dance at a club, or sing duets, or indulge in a large bowl of ice cream, and we can do much of this even if we or our co-participants lack some of the cognitive capacities typical of mature humans. This view of personhood is thus more expansive than the Lockean view (potentially answering the fetus and vegetable objections – a point to which I will return later).

There are complications for the view which must be acknowledged. It may seem at once too expansive and too narrow in its characterization of persons. The sense that the view is too inclusive arises because some of the interactions I describe as part of a person life are interactions we also have with animals that are generally agreed to be non-persons. It is common, for instance, for people to play with and talk to their pets, dress them in clothes, and give them beds to sleep in in much the same way people do with human infants. PLV's characterization of persons may seem too narrow because the kinds of interactions I have used to illustrate the wide range of activities in a person life are peculiar to lives lived in particular places at a particular time. The person-specific interactions with infants and those with dementia described earlier, for instance, do not occur in every culture or in every era. And even where such interactions are part of a culture, not every infant or dementia patient living within that culture will necessarily be treated this way. Some of those who we wish to count as persons, it seems, will be excluded by the view.

I offer some preliminary responses to these worries, which will lead in short order to the worries about conventionalism. As I speak to those worries, these brief initial responses will be developed in somewhat more detail. Starting with the observation that some people interact with their pets as they

might be an infant or small child, it is worth stressing that PLV embraces continuities between our interactions with human persons and our interactions with other animals. It is important to emphasize also, however, that there are a great many interactions specific to other people, including those who are not yet Lockean persons (e.g., human infants), and those who have ceased to be (e.g., dementia patients). Moreover, even interactions that are shared with both persons and non-persons take a different form with those we see as persons than those we see as non-persons. The ways we play with, instruct, feed, and clothe human infants are different from the ways in which we engage in these activities with non-human animals. Much of the difference resides in the fact that the way we treat human infants anticipates and supports their development into beings that would count as Lockean persons, while the way we treat beloved pets, even those seen as members of the family, anticipates and supports a different developmental trajectory.³

There may, of course, be cases of individuals who do not make this distinction, treating their pampered poodles, for instance, exactly as others would treat a human infant. Sadly, there are also instances in which individuals treat human infants as if they were non-human animals. According to PLV, however, these outlier cases play no role in determining facts about personhood. What it sees as critical to constituting personhood on this view are the practices inscribed in cultural and social institutions. In the present-day U.S., for instance, no matter how much someone sees her poodle as a person, the poodle will not be counted in the census, or provided an unemployment check, or enrolled in the public schools. On the other side, parents who feed their human children from bowls on the floor, leave them locked up alone at home all day, or do not send them to school do not make these children non-persons. Instead, the parents are viewed as monsters and the courts will intervene to take their children away from them.

This, then, is the general idea of person space. Laws, institutions, and cultural practice include implicit and explicit norms about the ways in persons interact with one another, creating a space of interpersonal interaction. Being included in this system of interactions is a constitutive element of living the life of a person, and so of being one. Once this feature of the view is stated explicitly, however, the worry about conventionalism is a natural one. If being a person requires inclusion in the space of interpersonal interactions, which amounts to being seen and treated as a person in accordance with the norms of the culture, it seems that a culture would have the power to determine which things are and which things are not persons, and this may seem both metaphysically and morally suspect. This is especially true given the concern about the narrowness of the view's characterization of personhood mentioned above and stemming from the dramatic differences over time and across cultures with respect to social and cultural institutions and practices concerning interpersonal interactions.

There are a variety of ways in which this worry might be developed, but perhaps one of the clearest is to focus on the status of adult humans who

possess the capacities that define personhood on the Lockean account. If personhood is fully or partially constituted by sociocultural norms and practices, it seems possible for such individuals to be non-persons in some cultures and persons in others. If being a person is a matter of being counted in the census or having the standing to bring a lawsuit or being inducted into a religious community or educated, history is replete with examples of cultures in which adult humans endowed with reason, reflection, and self-consciousness are excluded from personhood on the basis of race, sex, ethnicity, country of origin, religion, or similarly arbitrary reasons. Most cultures remain selective about which humans enjoy these person-specific privileges, and this is arguably, to some extent, a feature of all human cultures. If “person” is to name a kind, however, it is not obvious from a metaphysical perspective how something could be a person in one culture and not in another based solely on how they are treated by others. Ethically, there is a worry that the view implies that a group of humans who are oppressed or enslaved or stripped of basic rights because of race, creed, or origin have no basis for complaint because they are only persons if the culture sees them as such. These would, indeed, be troubling implications, but fortunately not ones which follow from PLV.

3 An anthropological account

To see why PLV need not be objectionably conventionalist, it is important to appreciate two things. First, the degree to which the conception of personhood it employs differs from the Lockean conception, and, second, the extent to which it sees sociocultural practices and institutions as constrained by natural fact.

There is a long tradition in philosophy of viewing the concept of personhood as inextricably bound up with a particular moral status and/or deservingness of respect. On this view, to be a person is to be worthy of a particular kind of treatment, and to see or treat someone as a person is to treat them with respect or grant them rights. It is a consequence of the way in which PLV expands the conception of personhood offered by Locke, and the range of interactions that count as person specific, however that this connection is not as tight on this view. To see and treat others as persons, according to PLV, is to see them as beings of the sort that are, typically, able to govern their lives by social and cultural institutions and practices. This does not directly imply that all such beings deserve equitable treatment or respect, and it decidedly does not imply that seeing and treating someone as a person necessarily involves treating them well or with respect.

It is, as already mentioned, part of the structure of many cultures to distinguish different levels or classes of humans and to codify different kinds of treatment for these different groups. According to PLV, this is to recognize different classes or levels of *persons* who are, according to local social norms, legitimately treated differently. It may be codified in law or practice, for instance, that kings or nobles deserve more respect than commoners and peasants. This does

not imply, as I am understanding personhood, viewing the commoners and peasants as non-persons. They are still beings which can be talked to and given instructions. They still need money and clothes. They can be reproduced with. More to the point, there are clear rules about their social place and what social goods are and are not available to them. In this way, they are given a clear place in the space of interpersonal interactions encoded in cultural practices and institutions, even if it is not as desirable a position as others.

A society can thus recognize classes of persons (as I am using the term) that it is permissible to mistreat in a variety of ways. Indeed, many of the most egregious forms of mistreatment are possible only when there is some fundamental level at which those mistreated are seen as fundamentally of the same kind as those dishing out the mistreatment. When humans degrade other humans by making them defile something they take to be sacred, for instance, or appear naked in front of others, or when they prevent them from voting they are, in doing so, recognizing them as persons. Only some kinds of beings can be tormented or controlled in these particular ways, beings who are capable of holding things sacred or being ashamed of their nakedness or who can understand what voting is and would be capable of doing so if not actively prevented. Any human social arrangement that oppresses other humans without recognizing them as persons in this fundamental sense (i.e., as beings that can conform themselves to the rules of social institutions and who might also aim to change them) would not last very long. Such a serious underestimation of the capacities of those one seeks to control is hardly likely to lead to success at controlling them. There may, of course, be very good arguments that it is morally indefensible to institute structures and practices that treat some groups of human persons worse than others. I happen to believe that there are. This is compatible with the claim that someone can be treated as a person without being treated as they should be. Indeed, the moral indefensibility of mistreatment is often accentuated by the fact that it takes place despite the recognition of what is fundamentally in common between those in a privileged position and those mistreated.

Once this aspect of the view is made clear, it may seem that PLV collapses back into a traditional psychological approach, according to which what makes someone a person is possessing the Lockean forensic capacities. PLV, it might be argued, merely emphasizes the role these capacities play in constructing and conforming to social institutions rather than those they play in individual interactions. This is not quite right, however. One critical difference between the traditional psychological view and PLV is the more diachronic, developmental perspective of the latter. PLV looks at a life trajectory rather than a slice of time and so, as described earlier, personhood consists not just in the possession of capabilities at a moment, but in being of a sort that typically develops these capacities at some point in one's existence. In human cultures, as already noted, this means that others are interacted with as persons before developing the Lockean capacities and can continue to be interacted with in this way after they lose them.

Another critical difference between the two approaches is that according to PLV, inclusion in person space is not a decision that is made on a case-by-case basis. We do not, that is, look at individual humans and decide whether they are to be brought into the group which is governed by person-related practices. PLV claims, in fact, that in human cultures and social organizations all humans are automatically seen and treated as persons (in the sense just described). This is where the anthropological features of PLV come in. Human cultural and social organization is recognized as having co-evolved with human biology, which means that as sophisticated and varied as human social and cultural organization is, it is at bottom constrained by our human biological nature. Recognition of other humans as of the sort with the potential to participate in sociocultural institutions, and so as persons in the sense at work in PLV, is thus a deep and non-accidental feature of human social organization.

I will say more in support of this point later, but for the moment we can note that the inclusion of all humans in person space is to be expected given the expansion of relevant interactions and judgments that are relevant to personhood on this view. We interact with one another in our bodies. Singing, dancing, athletic contests, having children together, and holding hands in the movie theater are all person-specific interactions, and they depend as much upon how we are embodied as on any facts about reason and reflection. Recent work in embodied cognition and enactivism suggests, furthermore, that even what seem the most body-independent elements of our cognition in fact depend upon our embodiment and our social and natural environment. (See, for instance, Gallagher, 2019 or Hutto and Myin, 2017.) Given that all humans are similarly embodied, and that evolutionary processes would be expected to favor a tendency for our conspecifics to be especially salient to us, as other animals are to them, it is not surprising that human cultures include a recognition of a special status for other humans, even as they make discriminations among different groups of humans.

The contours of person space are thus set by biological facts, among others. While human history makes it clear that there is room for a great deal of variation within the constraints set by biology, PLV insists that at bottom, human sociocultural institutions and practices include other humans in person space as a matter of course, even if the form of that inclusion is not always fair or desirable. This claim, I recognize, requires more defense than I have given it here. For now, however, I only note that even if it is granted, it appears to lead directly to the second horn of the dilemma described at the outset. If it is the case that all humans are included in person space in virtue of being humans, then why do we not just say, as Olson urges, that we human persons are fundamentally human animals? PLV adds the claim that such animals live in a complex social world which is very important to them. It might appear, however, that it does not give us any reason to think that social interactions and practices, no matter how fundamental, play any role in constituting human persons as entities or in determining the conditions of their persistence.

Instead, it seems to concede that metaphysically speaking we are human animals since it is a human animal that lives in culture and undertakes these complex social activities. It remains to be shown how PLV can resist this conclusion.

4 Differences from animalism

The denial of the claim that PLV is in the end a form of Animalism rests on an insistence that this suggestion does not take seriously enough the fundamental way in which social functions are intertwined with biological ones, and the degree to which these interactions are implicated in maintaining the integrity and persistence of individual human persons. Here, again, the anthropological nature of the view becomes important. I mentioned earlier that at the evolutionary timescale social and psychological factors interact with biological ones, and so cultural and biological features of human persons co-evolve. This means not only that social organization is constrained by biology but, equally important, that human biology develops in response to social organization in an ongoing feedback loop (e.g., dental structure and digestive enzymes adapt to the institutions of agriculture; features of the immune system alter in response to the development of sanitation). The interaction between biological and social factors does not occur only over evolutionary time, but also within the span of a single life. The biological development and functioning of an organism depend upon social interactions within person space (e.g., the development of brain architecture in infants depends upon interactions with caregivers; social isolation leads to a variety of health problems).⁴ Biological and social factors thus interact both in the history of beings like us and in real time as we live our lives. Where Animalism sees the organism as the fundamental entity and views social and psychological factors as add-ons whose removal may affect how we value the relevant entity but do not change its nature, PLV sees human persons as fundamentally bio/psycho/social beings, with no one set of functions more fundamental than the others. Biological, psychological, and social features all play a role in determining the persistence conditions of beings like us on this view.

The philosophical discussion of personal identity, of course, begins from the idea that biological and psychological continuity can come apart, and clearly social continuity can also come apart from either of the other two forms. It is thus necessary to consider what PLV has to say about the persistence of human persons in such cases. The view does allow, as does Animalism, that an individual human person could survive the loss of typical psychological or social functions of human persons. This possibility does not, however, undermine the claim that these features are as fundamental as biological ones to maintaining the integrity of human persons at and over time. Seeing why this is so, however, depends upon taking a metaphysical perspective deeply different from that which is usually presupposed in discussions of personal identity.

Olson very helpfully lays out the way specific metaphysical assumptions function in the argument for Animalism (Olson, 1997, pp. 3–6; Olson, 2007). These assumptions are shared with traditional psychological approaches to personal identity and form the background against which the debate between the two positions unfolds. Both sides of this dispute start by assuming that individual entities are given as members of a kind and that as such they have traits which they could not lose without ceasing to exist. Animalists hold that we are essentially human animals and could not continue except as human animals. Psychological theorists hold that we are experiencing subjects and can only continue as such. Confronted with a human person, who is both an animal and a subject, then, these theorists see a need to determine which is the kind to which it fundamentally belongs, and so which trait it could not lose without ceasing to exist.⁵ Animalists add the further assumption that “human animal” is a paradigmatic kind of substance. Anything that is an animal, in other words, must be essentially an animal. These assumptions, if accepted, do make Animalism almost irresistible.⁶

PLV, however, proceeds from a different metaphysical model. This view offers a cluster account according to which the different kinds of functions that are typically involved in maintaining human persons are all part of their fundamental nature, with no one of these kinds of continuity being in itself necessary for persistence. Perhaps the easiest way to understand this claim is via an analogy with a similar kind of view about the biological persistence of organisms, which has been offered by Winston Chiong. Chiong enters the argument between those who defend a circulatory-respiratory criterion of biological death and those who support a whole brain death criterion. Rather than trying to determine which of these is the correct criterion of death, Chiong suggests, we should instead think of life and death as cluster concepts, such that

the property of being alive (like the property of being a language or a game) involves a cluster of characteristics – none of which is itself necessary and sufficient for an organism to be alive, but all of which contribute to an organism’s being alive and tend to reinforce one another in paradigm cases.

(Chiong, 2005, p. 25)

A standard and plausible general definition of death for organisms is “the irreversible cessation of functioning of the organism as a whole”, and biology suggests that there are a variety of ways in which organisms can continue to function as wholes, at least if they are given the proper support.

In the paradigm case, organic life involves the functioning of all major organs and systems. The organism can be maintained through the loss of any one of these systems, however, so long as the compromised functions are compensated for in some other way. There is a great deal of redundancy built

into the human organism, and artificial life support systems, medications, dialysis, insulin injections, antibiotics, and antivirals can all do some of the work that natural systems are no longer capable of carrying out. Where these artificial compensations are not possible, the organism may, depending about the centrality of the function lost, continue to function for some time after the failure of almost any internal or natural system. The traditional approach to understanding death tempts us to define "life" in terms of that last function whose cessation leads to the disintegration of the organism as a whole. But "lastness" may be merely chronological and not consistent from case to case. There are different ways of dying, and they involve different kinds of specific last moments. This means that there are a variety of functions whose cessation might (or might not) undermine the integrity of an organism, depending on which other functions are still occurring, and what kinds of artificial support are available.

PLV takes a similar view except instead of including only biological functions, it includes also the psychological and social functions that, in paradigmatic cases, work together with the biological to maintain the form and functioning of the individual human person. Particular cases will, of course, deviate to some degree or another from paradigmatic ones. The question of how much deviation can occur without the integrity of the individual being lost and the individual human person ceasing to exist is going to be a complicated one whose answer can vary with context in the ways Chiong describes with respect to biological functioning. The key point, however, is that according to the cluster picture of what constitutes human persons employed by PLV, we are not forced to determine which of biological, psychological, and social continuity is *the* continuity that cannot be lost without the person ceasing to exist. Indeed, to try to do so is to misunderstand the source of the integrity of human persons, which consists in dynamic interaction among these features. Understood in this way, the loss of any of these forms of continuity is in principle survivable, although often in a diminished form.

PLV can thus allow that human persons can survive in a persistent vegetative state (PVS), but not only because a single living organism persists throughout this alteration. Myriad practical relations also continue, and so a PVS patient maintains a location person space. There is one sense, of course, in which there are no meaningful *interactions* with someone in such a state but, as mentioned before, someone in a PVS also does not play the sociocultural role of a piece of meat or medical waste. Such individuals are clothed and cared for, their spouses are still married, and so on. Admittedly, this will not be true in every culture or social organization. But in societies where this is not true, the biological trajectory of PVS patients will also be drastically different; without such interventions a PVS patient will soon be biologically dead.

Importantly, however, PLV also has the consequence that a human person could at least in principle survive being transplanted into another human

animal by one of the myriad science fictional technologies envisioned in the personal identity literature. We need, of course, to be careful about how we use thought experiments that are so far from the world as we know it, and to be cognizant of what we can and cannot learn from them. I see cases of this sort as an imaginative exercise that invites us to predict how social practices would develop under the circumstances described. What we are interested in determining is whether, under the scenarios envisioned, we are able to identify a single individual that occupies an ongoing place in person space, and if so, what the trajectory of that individual looks like. With this understanding, I believe we can describe very specific kinds of brain transplants and “body swaps” in which we are overwhelmingly inclined to judge, from the point of view of our current practices, that were they to occur we would track a person into a new animal. In such cases, PLV has the implication that the person would survive in a new body, although there would likely be many accommodations to be made.

I realize that these claims require more defense and explanation than has been given here, and that there are multitudes of puzzle cases and contradictions that could be raised to cause trouble for PLV. While it is possible to answer some of these, I think we need also to accept that this will be the case with any developed account of personal identity, including the traditional psychological approach and Animalism. I hope, however, to have at least provided a basic understanding of what PLV says and how it can steer between the two horns of the dilemma described at the outset. It is not conventionalist, because it allows that facts about personal identity are constrained by biological facts about human persons. It is not Animalism, because facts about identity are equally dependent upon and constrained by social facts, and this is because the connection between human social organization and human biology is not superficial or accidental, but deep.

The lesson from *Wakefield* thus applies. There can only be persons because there are social systems and practices, and individual persons require these to maintain them as single continuing beings. Stepping out of the system of human interactions will probably not, on the cluster view, lead to the complete cessation of a person's life. An adult, enculturated human who is shut off from other humans will probably have internalized enough socialization to continue, although not without the well-documented effects of social isolation. The question of whether someone continues over time and under what conditions is thus a complex one, which is sensitive to biological, psychological, and social facts. We do not, according to PLV, make it so that an individual at t^2 is the same individual as one at t^1 just by saying or thinking that they are, but the practices of interacting with and tracking others day in and day out have a profound impact on how human persons develop and function at all levels, including biologically. These practices constitute the person space in which we live the lives of human persons and are part of what makes us who and what we are.⁷

Notes

- 1 This view is developed in detail in Schechtman, 2014.
- 2 The view does allow for non-human persons, but the explanation of how, and of their connection to human persons, is somewhat complex and would take us too far afield. For more details about this feature of the view see Schechtman, 2014, pp. 131–137.
- 3 Or, for those who have ceased to be persons, we treat them as someone who has a history of having participated actively in culture in a way that pets have not.
- 4 For discussion of the co-evolution of culture and biology see, e.g., Donald, 2001 or Henrich, 2016.
- 5 It is worth noting that Locke's own metaphysics are quite different from this, and so he would not face the same challenge from Animalists that present-day psychological theorists do.
- 6 However some who share these basic assumptions have resisted. See, e.g., Baker, 2000.
- 7 I am deeply grateful to the participants in the International Conference on "The Unity of a Person: Ontology – Subjectivity – Normativity" sponsored by German Research Foundation and held at Ludwig-Maximilians Universität Munich in April of 2019 for their feedback on an early version of this paper, with special thanks to Jörg Noller.

Bibliography

- Baker, L. R., 2000. *Persons and Bodies: a Constitution View*. Cambridge: Cambridge University Press.
- Carter, R., 1989. How to Change Your Mind. *Canadian Journal of Philosophy*, 19, pp. 1–14.
- Chiong, W., 2005. Brain Death without Definitions. *Hastings Center Report*, 35(6), pp. 20–30.
- Donald, M., 2001. *A Mind so Rare: the Evolution of Human Consciousness*. New York: Norton.
- Gallagher, S., 2019. *Enactivist Interventions: Rethinking the Mind*. Oxford: Oxford University Press.
- Hawthorne, N., 1982. Wakefield. In: R. H. Pearce, ed. *Hawthorne: Tales and Sketches*. New York: Library of America, pp. 290–298.
- Henrich, J., 2016. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating our Species, and Making Us Smarter*. Princeton, NJ: Princeton University Press.
- Hutto, D. & Myin, E., 2017. *Evolving Enactivism: Basic Minds Meet Content*. Cambridge: The MIT Press.
- Inwagen, P. v., 1990. *Material Beings*. Ithaca, NY: Cornell University Press.
- Lewis, D., 1983. *Philosophical Papers, Vol. 1*. New York: Oxford University Press.
- Locke, J., 1975. *An Essay Concerning Human Understanding*. Oxford: Clarendon Press.
- Mackie, D., 1999. Personal Identity and Dead People. *Philosophical Studies*, (95), pp. 219–242.
- Noonan, H., 2003. *Personal Identity*. London: Routledge.
- Olson, E., 1997. *The Human Animal: Personal Identity without Psychology*. New York: Oxford University Press.
- Olson, E., 1999. Reply to Lynne Rudder Baker. *Philosophy and Phenomenological Research*, 59(1), pp. 161–166.

- Olson, E., 2007. *What Are We?: A Study in Personal Ontology*. New York: Oxford University Press.
- Parfit, D., 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Perry, J., 1972. Can the Self Divide?. *Journal of Philosophy*, 69(16), pp. 463–488.
- Schechtman, M., 2014. *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford: Oxford University Press.
- Shoemaker, S., 1984. Personal Identity: A Materialist's Account. In: S. Shoemaker & R. Swinburne, eds. *Personal Identity (Great Debates in Philosophy)*. Oxford: Blackwell, pp. 67–138.
- Snowdon, P., 2014. *Persons, Animals, Ourselves*. Oxford: Oxford University Press.

10 Personal identity and responsibility

Kathleen Wallace

1 Introduction: framing the issue of identity and responsibility

Many discussions of responsibility focus on questions about freedom or free will – is an agent free, did she choose or could she have done otherwise? The thought is that unless the agent is free, or could have chosen otherwise, then the agent isn't responsible for their action; the agent might be the or a cause of some outcome, but isn't responsible for it. For example, suppose a pedestrian suddenly jumps into the roadway directly in front of a vehicle (and the driver is observing all the rules of the road, speed limit, and so on) such that it is impossible for the driver to brake or swerve in time to avoid hitting the pedestrian. The driver may be causally implicated but would not be blamed or held accountable for hitting the pedestrian.

A related, but distinct approach turns on discussions of intent. Here the thought is that if the agent doesn't intend a particular action or outcome, then the agent is not or is less responsible for it. In the law, intent and premeditation have been considered relevant to determining the severity of a crime and the appropriate blame and punishment to assign.

One way of conceptualizing responsibility is to employ a distinction made by Watson (1975, 1996) between attributability whereby the agent is the subject of appraisal and deserving of (moral) praise or blame in virtue of their state of mind or adopted end and accountability whereby the agent is held answerable in some way by moral and social practices for their conduct.¹ In this chapter, I will employ these concepts but not analyze them per se. Rather, I am interested in an underlying issue concerning the identity of a self that changes over time and that is the proper target of responsibility assignment.

Responsibility as it pertains to questions about personal identity was a concern of Locke's who framed the issue in terms of the criteria for identifying the right person, the one who performed some action. This is a philosophical problem because persons change over time, and yet identity seems to require exact similarity. If persons change over time, they are not qualitatively the same from one time to the next. If identity depends on qualitative similarity, then the question arises, is someone the same person and responsible for prior deeds?

Locke argued that what matters for *personal* identity and thus, responsibility attribution is psychological continuity, and specifically personal memory. This criterion was criticized by Reid (1895) and others. There is an extensive, and fraught, literature in philosophy from Locke through contemporary psychological theories of personal identity on what a criterion for psychological continuity should be *and* whether it establishes identity. Parfit (1986) famously denied that it does, but argued that continuity was sufficient for all the things that matter including one's own survival. I am not going to review this literature, but will refer to some of the concepts as pertinent to the account of persons and personal identity that I will lay out.

As Locke saw, identification underlies considerations of judgments of responsibility, in both of Watson's senses of attributability and accountability. Identification cuts across both, whether someone is praiseworthy or blameworthy or whether they should be "held to account."² In legal contexts, the accountability, practical consequences of responsibility judgments can be significant (e.g., punishment, such as imprisonment or even death, levying of fines, and so on). What I want to argue is that the psychological criterion is neither necessary nor sufficient for ensuring identification of the person who performed an action. Rather, we need a different account of the person and of personal identity in order to answer the identification question. I am not denying the importance of psychological considerations (capacity, intent, and so on) to the overall practices of responsibility attribution and accountability. I am just saying that those are secondary to the question concerning identification of the proper target. It is possible for the proper target to be identified, and yet not attach responsibility to the person on the grounds of incapacity, absence of intent, or presence of positive or mitigating states of mind or ends-adoption, or consideration of mitigating circumstances. But determination about whether it is appropriate to assign responsibility presupposes that we've got the right locus, that the proper target has been identified.

2 Identity and responsibility: some examples in the law

As a practical matter, identification of the proper target is an empirical matter of gathering the necessary information to trace an event to the right initiating causal agent. But, even in the law the question that troubled Locke might still arise. Could someone change enough such that they are no longer the same person? If someone has a conversion experience, or undergoes change through dementia or other illness such that they have no personal memory of their prior actions, are they the same person? On Lockean or psychological theories of persons – if a psychological criterion is not satisfied, the answer would be "No, it's not the same person." Conversely, when a psychological criterion is satisfied but other conditions such as brain transplants or body swaps obtain, the answer would be "Yes, it is the same person."³

A concrete example in the law given by Diamantis (2019) concerns someone who was convicted for murder, but who decades later has developed dementia (and other health issues). Diamantis argues that the law is mistaken in assuming that identity is a stable concept. Drawing on psychological theories of personal identity, Diamantis argues that personal identity changes over time. He calls this a characterological account of personal identity. He then argues that since a person's (characterological) identity changes, a person is not necessarily the same person as the person who committed a crime 30 years earlier, and therefore, that a person who was once guilty (e.g., at the time of conviction) could become innocent (e.g., as they age, and develop dementia or other debilitating condition) and should not continue to be punished. The argument here seems to be that the individual is no longer both (1) the person who was a proper target of blame (guilt, or the attributability notion of responsibility) and (2) the person who should be held to account. Diamantis later distinguishes between personal and criminal identity (p. 2041), citing Wallace (2019a). He states that he is interested only in *criminal* identity as what matters in the law not the identity of the whole self and that he will treat personal identity as reducing to criminal identity.

Another possible example in the law that implicitly concerns identity is a case involving a divorced couple and an agreement concerning disposition of frozen embryos. Matsumura (2014) analyzes the case, *A.Z. v. B.Z.*, 725 N.E. 2d (Massachusetts 2000), which involved a couple who created frozen embryos and at the time signed an agreement (which was repeatedly signed over several years) that if they separated at some future time, the embryos would become the property of the wife. When the couple did subsequently separate (and divorce) the husband objected to the wife's ownership and use of the embryos. The court upheld the husband's objection and concluded that the agreement was not binding.⁴

The facts in this particular case are complicated, and there were several factors that the court considered. But the comment made by the court and of interest to me was that enforcing the agreement would force the husband to become a parent against his current will and state of mind (*A.Z.*, 725 N.E.2d at 1057). The court did not outright suggest that the husband was a different person. But, this comment seems to make a distinction between a past self and a current self that is being coerced by a prior self to abide by an agreement, as if that agreement were made by someone else and for which the current self is not responsible. Matsumura (2014) calls this the "different selves rationale" (76).⁵ The example is interesting because it raises the issue of the relation between different temporal stages of a self and the extent to which a current stage can be identified as the target of responsibilities and obligations defined by a past stage of the self (via a precommitment device).

These examples illustrate that philosophical concerns about the relation between personal identity and responsibility are live practical issues. They also show some limitations of a psychological theory of persons. Psychological states are highly variable, memory is unreliable, and over the sheer

course of living, persons typically undergo many changes in their psychological states, belief paradigms, normative commitments, and so on. While consciousness and self-awareness are central to human life and to what is thought to be valuable and worthwhile in it, psychological theories are unstable, in my view, as a basis by themselves for identity. If we are going to be able to coherently address the connection between identity and responsibility, there needs to be a better account of the relation between phases of the self over time and whether, in the face of change, it makes sense to say that past and future phases of a self are different selves.⁶

3 Meanings of the term “identity”

It is helpful to distinguish between two different uses of the term “identity,” what I will call identity-i (for identification) and identity-c (for characterization).⁷

We sometimes use the term “identity” to mean *not* identity in the formal sense of one and the same thing, but to refer to a set of characteristics, “an identity,” e.g., as Irish, or Jewish, or Female, or Black, or Muslim, or non-Binary, or American, or Feminist, or Operaphile or Spouse,... or in the criminal case referred to above, as having a “criminal identity.” This “characterization” meaning, identity-c, is distinct from the formal identity meaning (being “one and the same thing”), identity-i. Selves can have multiple identities-c, and can change such “identities-c” independently of the issue of identity-i. The personal identity and identity politics literatures are rife with confusion on this point. My cumulative network model (CNM) of the self keeps the two meanings of identity distinct and gives an account of identity-i that allows for identification of the one self through change over time (through changes in identities-c).

In what follows I will lay out the CNM of the self, its account of personal identity over time, and how that bears on identification of the target for responsibility attribution. I developed this model in my book, *The Network Self: Relation, Process and Personal Identity* (Wallace 2019a).⁸ According to CNM, the self is a network of related traits *and* the self is a process. As such the self *is*, in part, constituted by its history. Memory of one’s history (or parts of one’s history) is also for most selves an important experiential feature of the self. But, absence of memory does not mean absence of one’s history as constitutive of the self. Memory may be important to one’s self-understanding and *sense* of identity. Responsibility practices may consider memory, and other psychological features, such as capacity, intent, and control, in determining attribution and distribution of responsibility. But, these are distinct from the matter of identification of the proper target of such considerations.

The identification question is both a backward looking issue, that is, looking to the past and asking whether a person is the correct target with respect to a past action, choice, or event, *and* forward looking, that is, what does it mean to take responsibility going forward. These aspects of responsibility are related insofar as both are temporal questions about responsibility.

4 The cumulative network self and personal identity

In the analytic philosophical literature on persons and personal identity, there are two main competing views, the psychological and the animalist view of persons. On the former, inspired by Lockean intuitions, a person or self is defined by consciousness. Identity over time, persistence, or continuity as that self or person is defined in terms of some psychological criteria and relations (such as memory, sameness of core personality or characterological traits, and the like).⁹ On the animalist view, a person or self is conceptualized as an organism, like other biological animals, and its identity over time, persistence, or continuity is defined in terms of the persisting integrity of the living organism.¹⁰

I have argued elsewhere that both of these are too narrow as a view of persons (Wallace 2019a, 2019b). They do not adequately account for the temporal structure of or for social relations as constitutive of persons. Consider *being* a parent or a spouse. That is not just a psychological state or attitude; nor is it merely a biological trait; it is a relationally constituted trait of a self. The psychological view also tends to ignore embodiment. Even when it concedes embodiment as in an “embodied persons view” (EPV),¹¹ *person* is still identified as the psychological features; it’s just that they have to be (embodied) in something. Both psychological and animalist views tend to have a container view of persons. Persons (psychologically defined) are in a body, or persons (on an animalist theory) are a body bounded by the boundaries of the bodily organism. But social relations are not in a body, and if they are constitutive of selves, as I argue they are, then a different model of the self is needed.

My theory conceptualizes the self as a process (cumulative) and as a network of interrelated psychological, biological *and* social, cultural, semantic, political traits. My CNM is largely in sympathy with Schechtman’s person life view (PLV) of the self (Schechtman 2014) and with Schechtman’s argument that practical considerations should be incorporated into a theory of selves as they actually are and as they are the target of forensic concerns (Schechtman 2008, 2014).

I do, however, distinguish my view from Schechtman’s in several respects:

- 1 CNM explicitly conceptualizes the self as a network and as a process. Schechtman’s PLV acknowledges social determinants, but does not conceptualize the self as a network; it also recognizes developmental aspects of a self, but does not conceptualize it as a process.
- 2 While PLV is not a straight up psychological view of the self, and Schechtman is at pains to incorporate embodiment and the variety of ways in which selves function, have social determinants, and live their lives, her view tends toward what I think of as a narrative view of the self. That is, there is a tendency to conceptualize the self in terms of meaning giving activities, both individually and socially. I think of this as a high-level function of selves. I have an account of the reflexive capability of a self that enables it to engage in meaning giving (and autonomous, or

self-governing) activity.¹² However, on CNM a self-process is broadly constituted and may persist as that particular self, albeit much diminished, even when the self loses many of the abilities that are taken to be important to personhood (such as self-awareness, memory, self-narration, and meaning bestowal).

According to CNM the network is also a process and thus is not a static, but a dynamic, evolving structure or system of relations (that may have normative dimensions, chosen or inherited¹³). A fictional illustration of the idea: imagine Lindsey who is mother, daughter, sibling, spouse, philosopher, feminist, resident of New York city, driver, US citizen, native English language speaker, Huntington disease (HD) carrier, left-handed, Irish-Catholic. The interrelatedness of the traits as a network can be represented graphically, as in Figures 10.1–10.3.¹⁴

In addition, traits may form hubs or clusters of traits, such that there is structure or organization to the network. There are also many relations between individual traits across “hubs” or clusters, such that traits are not “siloed” in hubs or clusters. Figure 10.2¹⁵ is an oversimplified representation, introduced to get at the idea of clustering.

Figures 10.1 and 10.2 represent the self at a cross-section of time. However, the self-network is also a process. It acquires and loses traits from one time to the next. Thus, Lindsey at 5 years old is a kindergartner and 42 inches tall; Lindsey at 35 is a philosophy professor and 68 inches tall. Traits are constituents of a self. (I use “trait” and “constituent” interchangeably.) While Lindsey

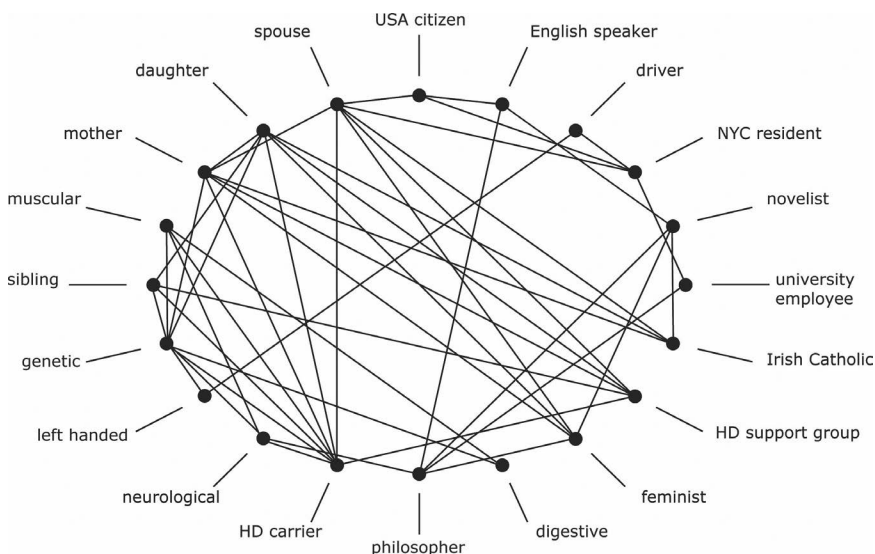


Figure 10.1 Network of a person's traits

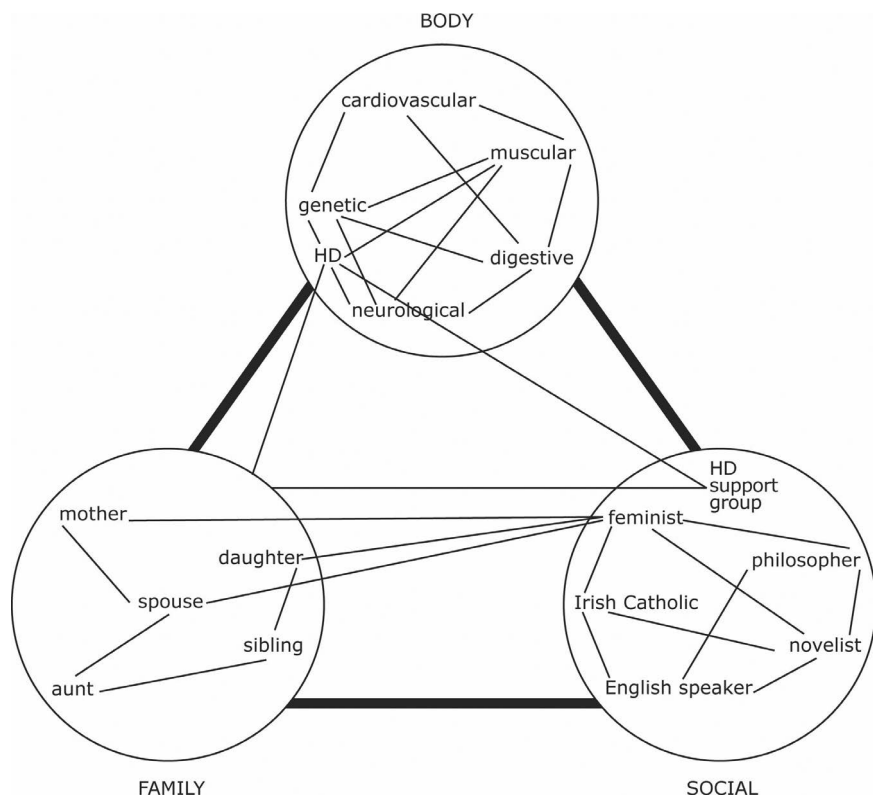


Figure 10.2 Clusters of a person's traits

at 35 is not a kindergartner or 42 inches tall, Lindsey at 35 is constituted (in part) as that self by her past; her past, including Lindsey at 5, is a trait of Lindsey at 35. A trait or constituent is a relevance condition of Lindsey being who Lindsey is (what I call her overall integrity). Lindsey's past is not literally present, but qua relevance condition of Lindsey is a constituent of Lindsey.

Figure 10.3¹⁶ attempts to represent this notion of the self as a cumulative process, representing Lindsey at three stages: Lindsey at 1, Lindsey at 5, and Lindsey at 35.

The idea is that Lindsey at thirty-five is the cumulative upshot of the whole process up to thirty-five, not just a discrete segment.¹⁷ Rather, each phase of the self has its past as a trait, represented by the nested boxes. Each phase is the cumulative upshot of the preceding stages of the self-process and is what it is in virtue of the preceding stages. This structural point is true of Lindsey at any stage.

Lindsey at any time is a cumulative upshot of the self-process. As such, the past is a constituent of the self-process being what it is; that's part of

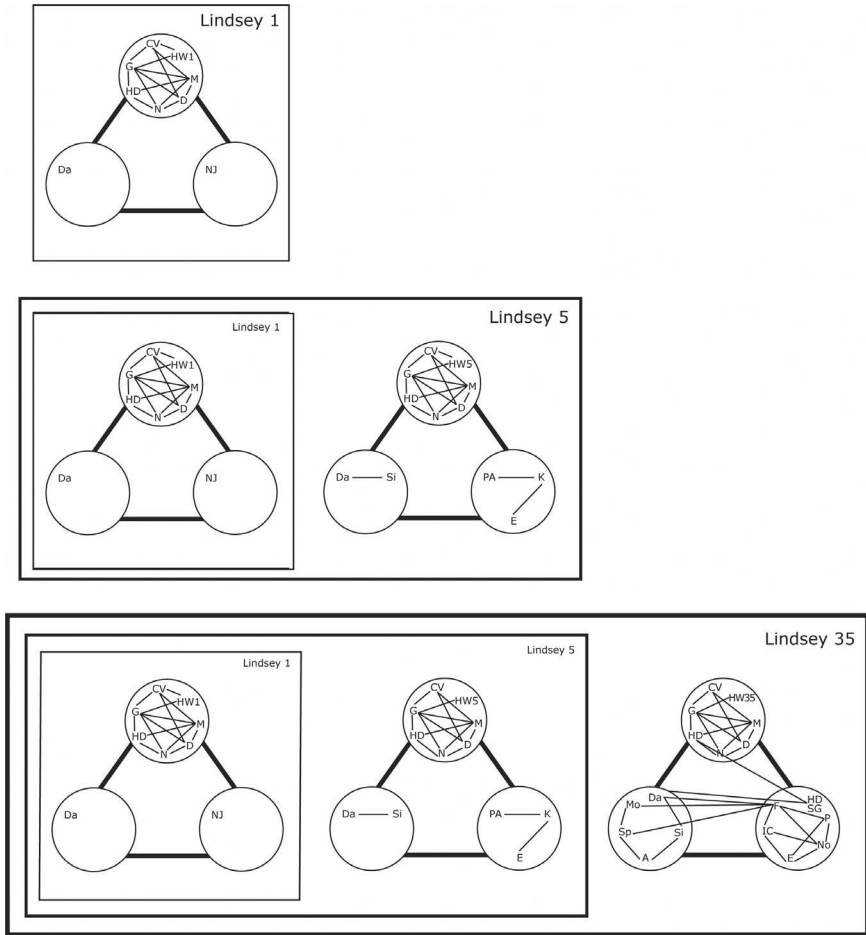


Figure 10.3 The self as a cumulative process

what it means to be a process. Not every constituent of Lindsey's past continues to be *specifically* relevant to who Lindsey is at the present. Some past traits or experiences may continue to be presently relevant, *qua specific*, and others are constituents of the cumulative past that itself is a constituent of the present self. For example, the genetic code may be presently relevant across many stages. But, that Lindsey at five is a resident of Pennsylvania (PA) may not be specifically relevant to Lindsey at thirty-five.¹⁸ But Lindsey at thirty-five is the cumulative upshot of the self-process. Thus, resident of PA at five is a constituent of her history, and that history or process is relevant as a cumulative whole to who she is now. On the other hand, suppose Lindsey at five suffers a trauma that, as that specific experience, continues

to shape and haunt her at thirty-five. Or, suppose Lindsey at sixteen is a student of an inspiring, formative teacher and that experience continues to shape Lindsey at thirty-five. Each of these – trauma experiencer, student of inspiring teacher – might be a specific trait that continues to be constitutive of Lindsey in the present in addition to being a constituent of the cumulative history.

Conceptualizing the self as a process and recognizing the past as a constituent of the cumulative self allows for change, even dramatic change, of one ongoing self-process. Rather than conceiving of a process as only a succession of stages (for comparison see note 15), I am suggesting that each stage (cumulative upshot) overlaps with the previous stage, or alternatively, each previous stage maps onto the subsequent stage (Wallace 2019a, 47ff). Since this is true at each stage (the cumulative upshot at any time), the whole processual past is a constituent at each stage, as the nested boxes in Figure 10.3 are meant to represent. The self's past need not be remembered to be relevant to who the self is in its current stage. Even when there are memory relations, much of the past may not be accessible to the consciousness of the self at subsequent stages. But that doesn't mean it isn't constitutive.¹⁹

A cumulative network is structured, is organized, by the relations among its strongly relevant traits. Exactly what is strongly relevant to what for any particular self being the unique self that it is is contingent. For instance, being a sibling may be a strongly relevant trait of one self, but weakly relevant of another.²⁰ (For the sake of simplicity, the figures do not distinguish between strongly and weakly relevant traits.) The interrelated unique organization of the cumulative network I call its integrity, by which I just mean its determinate nature.²¹ A self may also actively shape strongly relevant traits and thus, its integrity; the self is not merely a passive process.²² The integrity admits of change – it is not fixed, as indeed, *qua* process, it could not be.

The cumulative network does not admit of indefinite or just any change. I have argued elsewhere that the integrity depends on the persistence of some *cluster* of strongly relevant traits, but that there may be no single trait that is necessary or sufficient.²³ Suppose a person loses several limbs in a war or an accident, or loses memory due to a disease condition such as Alzheimer's. The integrity of the cumulative network, the self-process, persists, now in a diminished state; it maps onto the now-altered configuration of the network. That past is *its* history and continues to be constitutive of the cumulative network, even if the self-process persists with altered, diminished functionality. Contrast this with a person who is killed in a plane crash, their body blown apart and all that is recovered is a finger. The finger is a fragment of the cumulative network that had been that person. There is no longer a sufficient cluster of strongly relevant traits for the integrity of the person to "map on to." There is no self-process that persists. It is a trait of the finger that it had been a part of the body network (itself a sub-network) of the self; but that does not mean that the *self* persists. Rather, in this latter case, the self has ceased, while in the former two cases, the self persists *as that self*.

Selves can undergo significant change, imposed or chosen, such as a conversion experience or sex change, but not indefinite change without ceasing to be a or that self. Thus, according to CNM, personality changes, alterations in scope of functionality, biological and/or psychological, and cognitive changes by themselves do not entail cessation of one self or the beginning of a new one. A new state, even when it entails dramatic and sudden change (e.g., as a result of an accident), is still change *of* that self-process. As long as there is persistence of a sufficient cluster of strongly relevant traits (e.g., some combination of social, biological, physical, or psychological traits) that maps on to the subsequent cumulative stage – as in the example of limb loss, or memory loss, but not when all that remains is a finger – then the self persists, and is identifiable as that self. There may also be conditions such that a change or alteration is best described as aging or dying.

5 Identification

Earlier I distinguished between identity-i and identity-c, the identification and the characterization senses of identity.

Identity in the formal sense (what is often meant by identity-i) is a static notion but a process, by definition, is not static. Between phases, stages, or times of itself there cannot be identity in the sense of formal sameness since the self at different phases, stages, or times is different. Two problems that I think many approaches to personal identity have had are that they assume that each phase or stage *is* the self, rather than seeing that they are phases or stages *of the self* and there is a lack of recognition of the cumulative constitutedness of each phase. But on my approach, when we ask about the identity of, for example, the person in the courtroom – e.g., “is this the person who committed the crime?” – we are asking, not is the person in the courtroom identical with, meaning, “same as,” the person who committed the crime. Rather, we are asking an identification or numerical unity question: is this the person, the one who committed the crime? Alternatively, is the person in the courtroom the present summed up stage of the cumulative network who at an earlier stage committed the crime, who has that act as a constitutive element of its past? This is an awkward locution, but I think that’s the underlying structure. From one time to another the identification or identity-i question is a numerical unity question – is this “the one” the person who at t_n did action A and who at t_{n+1} is in the courtroom.

A spatial analogy may clarify the point. Consider the body – the arm and the leg are not identical to one another. But they do *belong* to one body. Colloquially, we’d say they are parts of one body and that (the body) is identical with itself (leaving aside temporal change of the body for the moment).

Analogously, there is a whole – a process – of which phases or stages of the self are constituents, or to which phases or stages of the self belong. Phases or stages are not identical with one another; the five-year-old (phase of the)

self is not identical with the thirty-five-year-old (phase of the) self. But each phase belongs to the process, and the cumulative self (the self-process of which each is a phase) is, at any time, identical with itself.

The spatial analogy is limited because it doesn't capture the cumulative constitutedness of the self at any time.²⁴ Each phase of the self is the cumulative upshot of what it has been and thus includes as a trait of itself, its past. Therefore, qua constituent of the past of the self, a prior trait may overlap, so to speak, with the current phase. Moreover, traits and clusters of traits may overlap many phases. For instance, a biological trait (e.g., Lindsey's DNA) may overlap all phases of the self, or a social trait may overlap many phases (e.g., Lindsey being a professor of philosophy overlaps phases from Lindsey at 30 to current Lindsey at 35). An action or choice may result in a trait ceasing to be presently constitutive, even though it remains a constituent of the self's past. For instance, the executing of a decision to get divorced entails that the self is no longer spouse of the divorced prior partner. However, that trait, spouse of *x*, is a constituent of the self's past, and the past is a constituent of the present self (expressed as now being the former spouse of *x*).

On this account, a self that undergoes considerable, even dramatic, change – changes in identities-*c* – is still the one self with those changes as constituents of its past. A change of identity-*c* does not entail that it is a different self process. The self is still itself, for what it is is not only how it is now, but as cumulative, it is also constituted by its past and is identifiable (identity-*i*) as that one self.

One might ask, what are *the* identity-*i* conditions of a process, of a cumulative network? On the cluster approach I am taking, I don't think we can articulate criteria for definitively demarcating when a self has ceased and when the self persists albeit drastically altered. There may be cases when psychological continuity is definitive, but there may be cases where psychological criteria are absent, but some other cluster of traits is sufficient. And, there may be borderline cases where the boundaries are not clear. I don't think there is any clear way around that problem.²⁵

6 Reconsidering the examples

Returning now to the examples mentioned earlier in this chapter. Consider the first one, the person who committed a crime and who many years later has developed dementia and has no memory of their prior deeds. Diamantis had suggested that the person is no longer the criminal self but has become an innocent self, not really the same self, and therefore is no longer guilty (attributability responsible) and not the proper target of punishment (accountability responsible). The suggestion is that the change in the self means that it is no longer the same self.

This line of thinking seems to conflate identity-*c* and identity-*i*. But, using the CNM framework, that the person no longer remembers their past does

not change the past as a constitutive trait of who they are, or that they are identified as the proper target for assignment of responsibility. The person does not cease to have been the perpetrator of the crime; does not cease to be the person serving, say, a life sentence; and so on. The person doesn't "become innocent" if that means that they are no longer constituted as the self who committed the crime. The self's past continues to be constitutive of the self. As that particular self-process it is correctly identified as the agent of that deed. At the same time, their loss of memory capacity is also now a trait of that self, and thus constitutes a change in the overall cumulative upshot that the self now is. Does that constitute a change in present constitution such that they are no longer "a criminal"? Alternatively, is the criminal history of the self strongly or weakly relevant to the responsibility that the altered self has in the present?

One way to address these questions would be to argue that they are not so much about identity as they are more appropriately answered through moral, social, legal considerations in exercising more leniency in responsibility practices. In the case in question, the argument might be that while the individual retains their "criminal identity," incapacity would be grounds for mitigating or ceasing punishment; in other cases, remorse or other changes in state of mind might be grounds for changes in appraisal and response. In the case of remorse, or regret, it only makes sense insofar as one recognizes oneself and is recognized by others as the one who did some prior action that one regrets or for which one feels remorse. Similarly with the experience of forgiveness – it depends on a recognizable continuity of the self; the wrong-doer must recognize him or herself as the one who needs to repudiate *his or her* own prior deed (Griswold 2007, 50). None of these changes mean that it is no longer one and "the same" self. Whatever changes in judgments of how and in what sense to assign responsibility, they only make sense if it is, indeed, the one self who is being reappraised or held to account.

Another approach, and the one that Diamantis advocates, is to argue that there is a change in the integrity of the self that warrants different treatment. In the loss of capacity case, the person has lost a trait that is so essential to being a criminal that they have lost their "criminal identity" (an identity-c) as a present constituting trait of the self; "criminal identity" would not overlap with the present phase of the now-dementia-afflicted self.

According to CNM, a self may cease having a trait (an identity-c) as a specific strongly relevant trait of the present cumulative upshot that the self is. Recall thirty-five-year-old Lindsey who is no longer a kindergartner (although she is constituted by her past in which five-year-old Lindsey is a kindergartner). Other traits may overlap many phases of the self. One example might be being a parent. Barring the death of the offspring, once one becomes a parent, being a parent is (normally) a persistent strongly relevant trait, an identity-c that overlaps all subsequent phases of the self. Its salience, its relevance may be stronger or weaker at different phases. How one is a parent,

its relevance in the network, and how it is constitutive of the self may alter extensively through the phases of the self-process, and one's responsibilities as a parent may change through that process. It could also become a weakly relevant trait for some self who, for instance, renounces all parental rights and ceases to have any relationship with their offspring.

Returning to Diamantis's case, the argument would be that for the dementia-afflicted individual their criminal identity has ceased to be a current trait of the self (like Lindsey ceasing to be a kindergartner) or has become a weakly relevant present trait of the individual (like an adult parent with attenuated responsibilities for their offspring). Their crime is still part of their past, and they are the proper target of identification for assignment of responsibility for that deed. However, their cumulative upshot self has changed such that the determinants of the trait "being a criminal" are no longer (or only weakly) present, and it no longer makes sense to assign responsibility to that self. Perhaps a person who has truly reformed and reconstituted themselves and renounced their previous patterns of behavior and commitments has also renounced their "criminal identity." The basic argument is, if there is sufficient change in the cumulative upshot of who a person is such that some strongly relevant trait is no longer or becomes weakly relevant as a present constitutive element of the self, then responsibility assignment, too, may change. That this is possible would be consistent with the CNM analysis of numerical unity and identification of the self through change.

From a legal and practical point of view, I do not know which of these two approaches makes the most sense: leniency or change in identity-c. Diamantis wants to argue for the latter that at least certain kinds of change suggest that criminal identity may no longer be retained as a present constituting trait of a cumulative self. In the case of genuine self-reform, such an approach might encourage less stigmatization and help those who have reformed to move forward, while at the same time reduce unduly harsh treatments where they may seem unwarranted (one of Diamantis's stated concerns). On the other hand, there may be traits from one's past that continue to be specifically relevant to who one is in the present and are not easily shed. While a trait can become weakly relevant to a person's overall cumulative upshot, a more fine-grained analysis of the determinants of a specific trait such as criminal identity would be needed. Such an analysis would include consideration of whether retention of cognitive capacity and memory is necessary for retaining that *as a present constituting trait* or whether conviction and awareness at the time of conviction and sentencing are sufficient. Consideration of agential role (deliberate self-reform) versus imposed condition (dementia) to the change in self may also need to be considered to more fully address how responsibility assignment might alter. Here, I am only laying out a framework that provides for identity through change over time such that any change in responsibility assignment is still focused on the one self that has undergone change, not a different self that has replaced that self.

Now consider the other example cited earlier, in which a court commented that enforcing a prior agreement would force an ex-husband to become a parent against his current will and state of mind (*A.Z. v B.Z.*; Matsumura 2014; and see note 5). The agreement in question is a type of precommitment device, whereby a self sets a rule, guideline, or norm at time t_n for its conduct at time t_{n+1} . Precommitment devices are common enough in everyday life, from making promises to friends, to making loan agreements, contracts, and so on. They would not typically be thought of as either violating a self's autonomy (unless the precommitment or agreement itself was made under duress or coercion) or involving a change in identity-*c*. In fact, sometimes precommitment devices are insurance against temptations that are anticipated as possibly undermining the self's autonomy and "true" commitments.²⁶ That is not to say that they can't be revisited or renegotiated; for instance, someone who falls on hard times and cannot meet their financial obligations may renegotiate the terms of loan repayment or undergo bankruptcy proceedings. Or, a marriage that does not work out, in spite of best intentions and promises at the time of entering into it, and that is dissolved through divorce. Legally, the latter can be quite complicated and involve persisting obligations that arose from the marriage (e.g., alimony or other forms of spousal support, child support, eligibility for social security, and so on). With some precommitment devices, such as in those designed to insure autonomy in the face of temptations, the self may be in a temporarily difficult situation, but not experiencing a change in an identity-*c*. In other cases, for instance, divorce, in ceasing to be the spouse of someone, there is some change in the traits that are presently constitutive of a cumulative self.

The precommitment device in this instance, the agreement, envisioned the possibility of separation and what would be the disposition of the embryos under such a change in status. The court found that the husband's intentions at the time of signing the agreement were not sufficiently well-established and the terms of the agreement were too vague. The court also stated that even if the husband's intentions had been unambiguously expressed, it would not have enforced the agreement. Whatever one thinks of the court's reasoning, the philosophically interesting question is whether a change of mind upon actualization of the possibility envisioned by the agreement constitutes a change in the self such that enforcement of that agreement would amount to coercion, to "compelling someone to become a parent against his current will and state of mind."

It seems that the court is invoking *some* psychological criteria (the husband's change of mind post-divorce) as a basis for characterizing the agreement as "compelling him to become a parent against his will." However, even a psychological theory of persons need not come to such a conclusion. A theory that relies on some notion of memory would not regard a former spouse who has simply changed his mind as therefore a different person who is being coerced by a prior, past self. Matsumura (2014) analyzes this comment as

involving a mistaken assumption by the court that there is a distinction between present and future (and past) selves, such that it would be coercion for a present self that has had a change of heart to be bound by the agreement made by a prior, past self.

According to CNM, there can be some change in the present strongly relevant traits (and identity-c) of a self without that entailing a change in the identity of the self as the one self identifiable over time. This is because the self's past, its history, is a constituting trait of the self. In a case in which former spouses have divorced, that each is a spouse in a prior stage of their self process is still a constituent of their past. It is also possible that being a divorced former spouse may be strongly relevant in the present, for instance, through child care arrangements or because qua divorced the relatedness is transformed to an amicable friend relation, and each now has the trait "friend qua former spouse."

Suppose that at the time it was made, each spouse's intentions were clear and the agreement was specific. That would mean that the self envisioned itself as being constituted and obligated in its future. At the time of making the agreement, the self would have and be aware of itself as having a possible trait – namely, being separated (or divorced) at some future time. In envisioning that possibility for itself, the self would also assign itself the trait of committing to a particular arrangement of that future, in the example being considered, that ownership (and disposition) of the embryos would belong to the former wife. Fast forward, the envisioned possible alteration in the self is actualized (and presumably, the actualization comes about at least in part through each self's own choices and contributions to the dissolution of the marriage). In that actualized state, the self feels differently than it expected, has a change of mind, and regrets the precommitment. Whatever reasons there might be for a self to revisit such a precommitment, the actualization of a possibility that the self defined for itself is not a matter of the self being forced to do something. Rather, what this shows is that selves can actively contribute, so to speak, to the shape of their future self even if, as in this example, the self regrets its choices and seeks to alter the obligations defined by its choices. I am not arguing that a self is necessarily obligated to the terms of a precommitment device. My point is only that those obligations belong to the self and are not a matter of the self being forced or coerced.

The actualization of a possibility envisioned by a precommitment device is not a case of a different (the past) self acting on and against the will of a present self. Rather, each is a stage of the one self process, and in the case of a precommitment device, of a self actively shaping its own future and then retrospectively recognizing its own agency in shaping its possibilities and in actualizing some of them. The self is still the self who made the agreement, with the possible arrangements therein envisioned, even when upon actualization and a change in an aspect of a self's identity-c, the self reconsiders or seeks to alter conditions defined by its own precommitment device.

7 Conclusion

Responsibility and obligation are, by their very nature, temporal spanning concepts, and are both forward and backward looking. A self has agency in defining its own possible obligations and responsibilities as well as actualizations thereof. It can also be the case that changes are imposed on a self – for instance, the development of a disease state, such as dementia. I have not analyzed the extent to which agential contribution to change (such as deliberate effortful self-reform, or articulation of a precommitment device) versus imposed change (such as dementia) would affect whether obligations or responsibilities are retained. My point is that a current self is always, in part, the cumulative upshot of what it has been and to which obligations and responsibilities determined by its past belong.

What I have been proposing is that conceiving of the self as a cumulative network provides a richer framework for conceptualizing the identity of a self through change and for locating analyses of conditions under which obligation and responsibility are assigned and retained or not. As a process, the past of the self is constitutive of what the self is, and this constitutiveness is not merely a matter of memory or of other psychological states. Because the self is a process, it changes, but is identifiable as the one self insofar as there is a sufficient cluster of strongly relevant traits from one phase of the self to the next to sustain the numerical unity of that cumulative network. That the self has this unity over time is part of what makes sense not only of responsibility practices and precommitment devices as considered in the examples here, but also of experiences of regret, forgiveness, planning, and of the ways in which social relations, involvement, and commitments constitute a self over time.

Notes

- 1 Under legal strict liability, someone could be held accountable even when it is not their own conduct that is the cause.
- 2 “Assigning” responsibility, in whatever sense one means, depends on locating the right agent to whom responsibility attaches.
- 3 As in Locke’s prince-cobbler example, Locke (1975), Bk II, Ch. XXVII, section 15. Shoemaker, for example, argues that one can imagine a society “going in for a body-change.”

All of the social practices of the society presuppose that the procedure is person-preserving. The brain-state recipient is regarded as owning the property of the brain-state donor, as being married to the donor’s spouse, and as holding whatever offices, responsibilities, rights, obligations, etc., the brain-state donor held.

(Shoemaker 1984, 109)

Similarly, Quinton (1962) suggested that a six-year-old girl displaying Winston Churchill’s character would preserve *Churchill*. I have argued against these and think the conclusions drawn by psychological theories of personal identity in the dementia and the body swap cases are problematic (See Wallace 2019a,

Chapter 7). But, for now I will focus on the change in state of mind cases and not discuss brain transplants or body swaps.

- 4 A basic premise of contract law is that a contracting self (the self at time t_n) binds the future self (the self at time t_{n+1}).
- 5 Other preceding cases concerning similar disputes over frozen embryos held that agreements should be enforced, but these were in different states and did not constitute precedent for Massachusetts. (See *Davis v. Davis*, Tennessee 1992; *Kass v. Kass*, New York 1998.) Also, the facts in those cases may have been different than in *A.Z. v. B.Z.* In *A.Z. v. B.Z.*, the Massachusetts court argued that the state's public policy against forced procreation (and against enforcing contracts binding individuals to decisions about future family relations) was a very important consideration. The facts of this case are quite complicated and raise doubts about what exactly the former spouses were agreeing to (among the facts, are that the husband often signed blank consent forms which were subsequently filled in by the wife), although the Massachusetts court also stated that even if the agreement were unambiguous, it would still be a situation in which a self was being compelled against their will (*A.Z.*, 725 N.E. 2d, 1057).
- I cite the Massachusetts case and Matsumura's analysis of it only for the suggestion that there is a distinction between a present and future self such that they may not be the same self, not to comment on or analyze the merit of all the court's arguments or conclusion.
- 6 Limitations in psychological theories are also evident in other practical contexts (e.g., in treatment of Alzheimer's patients, persons with cognitive disabilities, amnesia cases) as well as in some of the fusion and fission thought experiments in the philosophical literature on personal identity. In Wallace (2019a) I discuss some of the former (71ff) and the latter (Chapter Four, 83–113). See also Wallace (2019b).
- 7 This distinction I borrow and adapt from Schechtman (1996, 2).
- 8 See Wallace (2019b) for an abbreviated account of the theory.
- 9 Proponents of psychological theories include S. Shoemaker (1984), Parfit (1986), Perry (1975), and many others.
- 10 Animalists would include Olson (1997, 2007), Snowdon (1990, 2014), Blatti (2012).
- 11 As expressed, for instance, by Parfit (2016).
- 12 See Wallace (2019a), Chapter 5, 114–141.
- 13 For example, as a language speaker, there are inherited norms of communication to follow if one wants to be understood; as a spouse, one may jointly set norms to govern the marriage.
- 14 Figure 1 is Figure 2.2, Wallace (2019a), 29.
- 15 Wallace (2019a), 32, Figure 2.5.
- 16 Wallace (2019a), 54, Figure 2.9.

This figure represents Lindsey as cumulative, structured subclusters of traits at a few temporal cross-sections or stages, Lindsey 1, Lindsey 5, Lindsey 35.

Abbreviations: A = Aunt; CV = cardiovascular system; Da = daughter; D = digestive system; E = English speaker; F = feminist; G = genetic structure; HD = Huntington Disease Carrier; HDG = member Huntington Disease Support Group; HW = height/weight at a phase; IC = Irish Catholic; K = Kindergartner; Mo = Mother; M = musculoskeletal system; N = neurological system; NJ = resident of New Jersey; No = Novelist; P = philosopher; PA = resident of Pennsylvania; Si = Sibling; Sp = Spouse.

The first box is not nested; it assumes that this is the starting point of the network of relations that is Lindsey and as such it doesn't yet have its own past qua network. A causal past that brings about the network may make it difficult to

conceptualize exactly when the network comes into existence. Here I stipulate that there is some initial time, birth, which is the beginning of Lindsey qua a distinctive self-network.

- 17 As would be the case with four-dimensionalist views of the self diagrammed as:

Lindsey 1	Lindsey 5	Lindsey 35
-----------	-----------	------------

- 18 One's prior residence could be presently relevant, for example, to one's citizenship and eligibility to be president of the United States, or to the ongoing character and experience of an immigrant.
- 19 Baylis (2017) has a poignant account of the ongoing integrity of her dementia-afflicted mother that illustrates this point and of the ways in which other (social) relations and traits continue to be constitutive of her mother.
- 20 In such a case, that it is strongly relevant at a previous phase would still be a constituent of the past of the self. For example, Mozart's sibling relation to his sister Nannerl seems to have been strongly relevant of Mozart as a young boy, but weakly relevant of the adult Mozart as musical composer.
- 21 I am not using the term in a value sense, as in the phrase "a person of integrity."
- 22 I give an account of this, developing what I call the concept of reflexive communication, in Wallace (2019a), Chapter 5, 114–141.
- 23 Wallace (2019a). I align my view with property cluster views in philosophy of biology (36–37), and deploy the cluster concept in the discussion of identity over time (70–72). Schechtman also appeals to the property cluster concept in her account of when a person life continues (Schechtman 2014, 8 and Chapter 6). For property cluster views in biology see Boyd (1991, 1999); Slater (2015).
- 24 This is one of the problems with four-dimensionalist accounts; they treat temporal parts as analogous to spatial parts. In a spatial analogy regarding body parts, parts (arm, leg) do not overlap with one another. The cumulative nature of temporal phases suggests that they are not parts in the same sense. Seibt's distinction between "being a part" and "being part of" may be useful in thinking about a phase "being part of" a process, as distinct from "being a part" (Seibt 2015). For four-dimensionalist views see, for example, Sider (2001), a stage theorist; Lewis (1976, 1986), who articulates what a temporal part is; Braddon-Mitchell and Miller (2004) who defend a version of temporal parts theory.
- 25 I analyze the implications of the cluster approach for identity over time in Wallace (2019a), 70–77 and Chapter 4, 83–113, on fusion and fission thought experiments.
- 26 Or, Ulysses-type arrangements, after Homer's *Odyssey* in which Odysseus (Latinized as "Ulysses") has his crew lash him to the mast with instructions to not free him, whatever his entreaties, as they pass the Sirens. See Elster (2000).

Bibliography

Baylis, F. (2017). Still Gloria: personal identity and dementia. *The International Journal of Feminist Approaches to Bioethics* 10, no. 1: 210–224.

Blatti, S. (2012). A new argument for animalism. *Analysis* 72: 685–690.

Boyd, R. (1991). Realism, anti-foundationalism and enthusiasm for natural kinds. *Philosophical Studies* 61: 127–148.

Boyd, E. (1999). Homeostasis, species and higher taxa. In R. Wilson (Ed.), *Species: New Interdisciplinary Essays* (pp. 141–186). Cambridge: MIT Press.

Braddon-Mitchell, D. and Miller, K. (2004). The loneliness of stages. *Analysis* 64, no. 3: 235–242.

- Diamantis, M. (2019). Limiting identity in criminal law. *Boston College Law Review* 60, no. 7: 2011–2099.
- Elster, J. (2000). *Ulysses Unbound: Studies in Rationality, Precommitment and Constraints*. Cambridge: Cambridge University Press.
- Griswold, C. (2007). *Forgiveness: A Philosophical Exploration*. Cambridge: Cambridge University Press.
- Lewis, D. (1976). Survival and identity. In A. Rorty (Ed.), *The Identities of Persons* (pp. 17–40). Berkeley: University of California Press. Reprinted with significant postscripts in Lewis, D. (1983). *Philosophical Papers*, volume I (pp. 55–77). Oxford: Oxford University Press.
- Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.
- Locke, J. (1975). *An Essay Concerning Human Understanding*, P.H. Nidditch (Ed.), Oxford: Clarendon Press.
- Matsumura, Kaiponanea. (2014). Binding future selves. *Louisiana Law Review* 75, 71–125.
- Olson, E. (1997). *The Human Animal: Personal Identity without Psychology*. New York: Oxford University Press.
- Olson, E. (2007). *What are We? A Study in Personal Ontology*. Oxford: Oxford University Press.
- Parfit, D. (1986). *Reasons and Persons*. Oxford: Clarendon Press of Oxford University Press.
- Parfit, D. (2016). We are not human beings. In S. Blatti & P. Snowdon (Eds.), *Animalism: New Essays on Persons, Animals and Identity* (pp. 31–49). Oxford: Oxford University Press.
- Perry, J. (1975). *Personal Identity*. Berkeley, CA: University of California Press.
- Quinton, A. (1962). The soul. *The Journal of Philosophy* 59(15), 393–409.
- Reid, Thomas. (1895). *Essays on the Intellectual Powers of Man* (1785), in *Philosophical Works of Thomas Reid*, W. Hamilton (Ed.), vol. 1; reprinted Hildesheim: George Olms.
- Schechtman, M. (1996). *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
- Schechtman, M. (2008). Personal continuation and a life worth having. In C. Mackenzie & K. Atkins (Eds.), *Practical Identity and Narrative Agency* (pp. 31–55). New York: Routledge.
- Schechtman, M. (2014). *Staying Alive: Personal Identity, Practical Concerns and the Unity of a Life*. Oxford: Oxford University Press.
- Seibt, J. (2015). Non-transitive parthood, leveled mereology, and the representation of emergent parts of processes. *Grazer Philosophische Studien* 91(1): 165–190.
- Shoemaker, S. (1984). Personal identity: a materialist account. In Sydney Shoemaker & Richard Swinburne (Eds.), *Personal Identity* (pp. 133–152). Oxford: Blackwell.
- Sider, Theodore. (2001). *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press.
- Slater, M. (2015). Natural kindness. *British Journal of Philosophy of Science* 66: 375–411.
- Snowdon, P.F. (1990). Persons, animals, and ourselves. In E. Gill (Ed.), *The Person and the Mind: Issues in Ancient and Modern Philosophy* (pp. 83–107). Oxford: Clarendon Press.
- Snowdon, P.F. (2014). *Persons, Animals, Ourselves*. Oxford: Oxford University Press.

- Wallace, K. (2019a). *The Network Self: Relation, Process, and Personal Identity*. New York & London: Routledge.
- Wallace, K. (2019b). A theory of the relational self: the cumulative network model. *Humana Mente: A Journal of Philosophical Studies* 12, no. 36: 189–220.
- Watson, G. (1975). Free agency. *Journal of Philosophy* 72: 205–220.
- Watson, G. (1996). Two faces of responsibility. *Philosophical Topics* 24: 227–248.

11 Collective intentionality, we-identity, and the role of narratives in the constitution of friendship

Felipe León¹

1 Introduction

What is the relationship between the first-person singular and the first-person plural perspectives? One influential way to tackle this question is in terms of the notion of collective intentionality, broadly understood as the power of human minds to be jointly directed to objects and states of affairs (Schweikard & Schmid, 2013, p. 1; Tollefsen, 2004). In the last decades, various proposals have been put forward in philosophy and neighbouring disciplines in order to explain how individual subjects can share mental states like intentions, beliefs, and emotions, and thereby be collectively directed to the world. Although collective intentionality is arguably a central aspect of the I-we relation, it is less often noted that the latter is not exhausted by the former. Another, less explored aspect of the I-we relation is collective or we-identity. In brief, whereas research on collective intentionality investigates the conditions under which a plurality of individual subjects may believe, intend, or feel something together, a key question about we-identity is ‘who are we?’.

To a first approximation, we-identities are the identities that individual subjects have as ‘one of us’, in virtue of their membership in specific social groups. This characterization doesn’t take us very far, though, because the notions of membership and social group can be understood in different ways. One can be a member of a social group irrespectively of whether one identifies with or cares about the group in question. Group membership may not even require that one knows about the group’s existence, let alone about one’s membership in it. For example, one can be a member of the group of people who have a genetic predisposition to increased serum calcium levels, in spite of one’s lack of knowledge about the existence of that group and one’s membership in it. In other cases, group membership presupposes a subjective appropriation. It is experienced from within as a feeling of belonging to the group (Taipale, 2019; Zahavi, 2019). Moreover, given the diversity of social groups, and the difficulties dividing them into simplified typologies (see Epstein, 2019), it is plausible that we-identities also vary depending on what kind of group is at issue. From two strangers who join forces to help someone struggling to get on a bus, to friendships, families, online communities,

institutions, nations, and ethnicities, the social world is partly constituted by a vast array of social groups that are plausibly linked to we-identities of various sorts.

The main aim of this chapter is to investigate a particular variety of we-identity, in which group membership is characterized by at least three features: (i) it is experienced from within, in the sense that being ‘one of us’ is—at least occasionally—an experientially salient feature; (ii) it is more robust and enduring than in transient groups formed on the spur of the moment, for example, in one single joint action; (iii) it is essential, insofar as the identity of the group doesn’t persist through changes of membership. By this I mean that if any particular member leaves the group, the identity of the group changes. Group membership characterized by these three features is present in a range of social formations that are a recognizable and central component of the human social fabric. In particular, such features are typically present in groups—and, minimally, dyads—of friends and romantic partners. Pairs of friends and romantic partners typically have affectively loaded attitudes of concern and care for one another. Such attitudes are experientially salient, diachronically extended, and other-focused, insofar as they target another particular person.

To illustrate, consider the following example of the phenomenon that, following Cocking and Kennet, I will call *companion friendship* (Cocking & Kennett, 1998). Alma and Milton are lifelong and close friends. They met at school, grew up together, and not only know each other very well, but have also shaped many of each other’s interests and values throughout the decades. They meet regularly, have their favourite cafés and restaurants, and enjoy spending time together. Intuitively, the way in which Alma and Milton are ‘one of us’ as friends is not something remote from their consciousnesses. Rather, as part of their friendship, they both experience care and affection for one another. Importantly, they don’t care for the other in view of the pleasure or benefits that they might individually extract from their relation, but rather for the other’s own sake.² Moreover, Alma’s and Milton’s we-identity as friends clearly differs from the way in which two strangers acting jointly may take themselves to be ‘one of us’. Two strangers who, suppose, spontaneously help someone to get on a bus may find it mutually advantageous to join forces to achieve that goal. But it is not important for them to be with one another as something inherently valuable and significant, as it happens for Alma and Milton.

Consider, further, that if either Alma or Milton leaves their group, the latter is thereby dissolved. Suppose that Milton moves to a different country and undergoes a religious conversion, and that this turns out to undermine the friendship with Alma. As time passes by, they progressively lose contact. In the meantime, Alma establishes a friendship with Marco, who after a long time comes to occupy a place nearly as special in Alma’s life as Milton did. In such a case, one wouldn’t be inclined to say that there has been a replacement of membership in an otherwise stable group,

but rather that a new group of friends constituted by Alma and Marco has been formed. The reason is that relations of friendship involve an appreciation of the particularity of another person that is in tension with taking the other to be replaceable by someone else (see Helm, 2017; Millgram, 1987, p. 363).

How can one go about the task of clarifying the status of we-identities like the one exemplified by the relationship between Alma and Milton? The main claim of this chapter is that one can do so in terms of a narrative conception of identity. The narrative approach is mostly known in connection with discussions about individual personal identity. It encompasses a family of proposals that have as common denominator the idea that narratives play a central role in shaping personhood (MacIntyre, 2007; Mackenzie & Atkins, 2008; Ricoeur, 1990; Schechtman, 1996). A relatively less explored topic in the literature is the relationship between narratives and the first-person plural perspective (Gallagher & Tollefsen, 2019; Tollefsen & Gallagher, 2017; see also Carr, 1986b). Building on this work, my main aim will be to outline a narrative approach applicable to the we-identities to be found in relations of interpersonal friendship like the one Alma and Milton have. On the proposal that I will develop, such we-identities are narratively constituted because they presuppose that participants communicate and conceive of the happenings of their lives—including some of their episodes of collective intentionality—in a unified fashion, as part of an unfolding trajectory that has some intrinsic value for them.

Although friendship is a multifaceted and complex phenomenon that can't be made fully justice within the scope of this chapter, I argue that the outlined proposal vindicates at least three characteristic features of companion friendships. The first one is that such friendships are historically constituted. They are backward- and forward-looking interpersonal relations that build on a shared history and sedimented habits. The second one is that friends stand in a communicative relation to one another. The third one is that friendship has a characteristic depth, in the sense that friends are in a position to literally shape aspects of each other's identity (see Cocking & Kennett, 1998, p. 504, 2000).³

The chapter is structured as follows. In Section 2, I elaborate on the distinction between collective intentionality and we-identity. I locate the relevant target phenomenon of we-identity in a broader context by introducing a distinction between three kinds of we-identities: *ephemeral*, *corporate*, and *personal*. Section 3 elaborates on the very notion of identity by drawing on discussions about self-identity. I propose that the distinction between identity in the sense of *re-identification* and in the sense of *characterization* (Schechtman, 1996) is applicable to personal we-identities, and that considerations analogous to the ones put forth in debates on self-identity support the claim that the characterizations question about personal we-identities is best answered in terms of a narrative theory of identity. In Section 4, I discuss Gallagher and Tollefsen's notion of we-narratives (Gallagher & Tollefsen, 2019). Since their notion of we-narratives is most readily applicable to corporate we-identities, I propose

some adjustments to make the notion of we-narratives applicable to personal we-identities exemplified by companion friendship. In Section 5, I elaborate on the notion of we-narrative applicable to friendship, and I argue that the adjusted notion accounts for the aforementioned features of the latter.

2 We-identities: ephemeral, corporate, personal

By and large, researchers working on collective intentionality tend to agree on a series of negative and positive constraints (see Schweikard & Schmid, 2013). The first negative constraint is that collective mental states are not a mere aggregation or summation of different individuals' mental states. For a plurality of individuals to have a collective mental state, it is not sufficient that they each have a mental state of the same type and with similar content. The second negative constraint is that collective mental states that are phenomenally conscious ought not to be attributed to a third, supra-individual entity, distinct from the involved individuals—a conscious group mind. There is nothing it is like to be a group, as distinguished from the first-person singular phenomenology of the group members. Whichever other properties groups may have, it is implausible to claim that they feel and experience like individual subjects do. The third constraint is threefold. It suggests three conditions that a convincing account of conscious collective intentionality has to meet: plurality, awareness, and integration. In short, conscious collective mental phenomena require more than one participant, each participant must have a conscious perspective on the world, and participants must be bound together in the right way (see Mathiesen, 2005).

Traditionally, debates on collective intentionality privileged specific mental state types, particularly intentions and beliefs, at the expense of other mental phenomena, like emotions and other affective states, that plausibly also come in the first-person plural. In the last years, there has been a lot of work that has amended this shortcoming, by supplementing and complexifying the palette of collective mental states and phenomena under scrutiny. Although much has been won from this strategy, it is important to realize its limits. What this strategy can provide is an investigation and clarification of a range of types of collective mental phenomena. But a fine-grained and taxonomically differentiated investigation of collective mental phenomena will not provide an account of how instantiations of those collective mental phenomena are interwoven and linked with each other in the unity of an identity. For comparison, consider personal identity. One feature that sets apart a person from an aggregate of mental states and character traits is that the former notion picks out a more or less coherent unit in which mental states and other psychological features are interwoven and made intelligible. This is what warrants attributions of what matters to a person, her responsibility, her take on a situation, and so on. Specific states and attitudes gain their intelligibility and significance as constituents of a psychological life in the context of the broader network of a *person's* mental states (Frankfurt, 1988; Wollheim, 1984).

Why should these considerations matter for a better understanding of the relationship between the first-person singular and the first-person plural perspectives? They matter because they suggest that no matter how exhaustive, fine-grained, and differentiated an account of collective intentionality is, that account will fall short of clarifying the status of the I-we relation if questions about identity are left out of the picture. If this diagnosis is correct, collective intentionality should be appraised as only one aspect of the I-we relation. The I-we relation is not only about intentionality, but also about identity. Elaborating on this point will help to better demarcate the target phenomenon of we-identity that will be under focus in this chapter.

Consider instances of collective intentionality in which individuals engage in short-lived cooperative activities like preparing a sauce together, taking a walk together, or dancing the tango together. Examples like these have been often discussed in the literature on collective intentionality. It has been argued that the performance of joint actions is typically accompanied by a distinctive sense of joint agency (Pacherie, 2011, 2014) or we-agency (Salmela & Nagatsu, 2017), and an experience of joint control (Tollefsen, 2014). And it seems indeed plausible that individuals who engage in single joint actions like carrying a table together or preparing a sauce together typically have some experientially salient ‘sense of us’ and of being ‘one of us’, insofar as they carry out a particular joint action together. However, one shouldn’t overlook the extent to which the sense of being ‘one of us’ in the performance of a joint action often depends on a pre-existing relationship between the involved agents. For example, compare a situation in which one is taking a walk together with a colleague whom one hasn’t met before, in the context of a walk-and-talk activity at a departmental seminar, and a situation in which one is taking a walk together with one’s romantic partner. Although, at some level of abstraction, the two situations have as a common denominator a sense of being ‘one of us’ while walking together, it is fairly clear that such sense will be different in the two cases. By focusing on joint action (and other instances of collective intentionality) in abstraction from pre-existing relationships that obtain between the involved agents, traditional theories of collective intentionality have tended to neglect questions pertaining to we-identity.

As a way of mapping the terrain, in a preliminary and non-exhaustive way, I propose to distinguish between three kinds of we-identities: *ephemeral*, *corporate*, and *personal*.⁴ *Ephemeral we-identities* are linked to “ephemeral groups” (Calabi, 2008), which arise and dissolve quickly. Such groups may be constituted in single episodes of triadic joint attention, small-scale joint action or emotional sharing. Paradigmatically, ephemeral we-identities take place in interactions between strangers. The only relevant pre-requisite to have an ephemeral we-identity in connection with the execution of a joint action is that individuals can cooperate. Ephemeral we-identities are quite pervasive in human sociality. One way to bring this out is by considering the role of communication in interpersonal understanding. The idea that interpersonal communication, in its different formats and varieties, is a widespread and

central form of social understanding is hard to deny. One proposal that is getting increasing traction amongst psychologists and philosophers is that joint cooperative (non-deceptive) communication is a joint action (Jankovic, 2018; Searle, 2010; Tomasello, 2008; see also Gallagher & Tollefsen, 2019, p. 213; Schmid, 2014, p. 11).⁵ If so, strangers who participate in short-lived joint communicative actions may have a sense of joint agency and joint control when they communicate, as it happens in other types of joint agency.

Whereas the notion of ephemeral we-identity arises from reflection about ephemeral groups, consideration of another type of social groups suggests a further notion of we-identity. This is the notion of what I call *corporate we-identity*. Corporate groups have been described in the literature as large-scale groups with a clear structure, an organization that is often hierarchical, and clearly stated decision-making procedures (French, 1984; Tollefsen, 2015, p. 47). Examples of such groups are business corporations, the US army, the Red Cross, and university faculties. Corporate groups are enduring groups that persist across changes of membership and have a legal status (List & Pettit, 2011; Pettit, 2014; Tollefsen, 2015).⁶ Such groups are usually contrasted with mere sets or aggregations, such as a crowd or the people standing at a bus stop (French, 1984, p. 12). Since the identity of a corporate groups persists across changes of individual membership, to have a we-identity in a corporate group involves seeing oneself as *representative* of a group that persists beyond one's membership in it. Although corporate we-identities don't require actual interpersonal interaction, when individuals relate to one another on the basis of a corporate we-identity, they relate to one another qua group members.

Might ephemeral and corporate we-identities have anything in common? I suggest that, in spite of their differences, they do. Neither ephemeral nor corporate we-identities require that individuals relate to one another *personally*. They may also take place if individuals relate to one another in an *impersonal* way. By 'impersonal' I don't mean disrespectful, objectifying, or de-humanizing, but rather the non-intimate way of relating to one another that is characteristic of interactions between strangers, passers-by or group representatives. Because of this shared feature, both ephemeral and corporate we-identities differ from *personal we-identities*, paradigmatically exemplified by dyadic companion friendships and romantic love relations. On the one hand, personal we-identities are more robust and enduring than ephemeral we-identities. On the other hand, they are less robust than corporate we-identities, given the lack of persistence of dyadic personal groups across changes of membership. Most critically, personal we-identities require appreciation of another particular person as a non-substitutable individual.

Recall the example of Alma and Milton, who have cultivated a lifelong friendship, and have grown together and shared their lives as friends. I think that the notion of personal we-identity is what we need to approach this example. Whereas the togetherness of Alma and Milton as friends surely differs from an ephemeral we-identity, it also differs from the robustness of corporate we-identities, insofar as their group doesn't have an identity that persists

if either of them leaves the group.⁷ More generally, what sets apart friendship and other relations in which private and intimate knowledge of a particular other plays a constitutive role is that they are distinctively personal relations.⁸

It is noteworthy that some recurrent examples of shared emotions and collective actions in the literature on collective intentionality involve individuals who have personal *we*-identities, in the sense I have just outlined. Consider Max Scheler's widely discussed example of emotional sharing: two parents are grieving in front of their dead child's body, and they feel their grief in common (Scheler, 2008, pp. 12–13; see León et al., 2019; Schmid, 2009). Some commentators have noted that this example “suggests pre-existing relations of marital love and marital life between the sharers of the feeling” (Konzelman-Ziv, 2009), as well as “a diachronic *narrative intimacy*” (Krueger, 2016, p. 270) comprising the memories, shared experiences, and associations peculiar to this particular family unit. And Scheler himself suggests that for one of the parents to join the other in the experience of grieving together, there must be in place “the highest form of love” (Scheler, 2008, p. 12).

Perhaps one reason why some paradigmatic examples of shared emotions, such as Scheler's, have proven challenging to tackle is that personal *we*-identities have not been brought properly into the picture. Be it as it may, my main focus in this chapter will be on personal *we*-identities, taking companion friendship as the guiding phenomenon under examination.

3 Re-identification, characterization, and narrative identity

This section takes the foregoing investigation of personal *we*-identities two steps forward by focusing on the very notion of identity and on the role of narratives in self-identity. Any attempt to shed light on the notion of identity will quickly realize that it is not univocal. It is helpful to consider here discussions about personal identity. Marya Schechtman has influentially argued that there is a distinction between two questions about personal identity: the *re-identification* question and the *characterization* question (Schechtman, 1996). The re-identification question inquires into the conditions that have to obtain for a person existing at one time to be numerically identical with a person existing at another time. The question here is how two independently definable moments or slices of a person can be said to be moments or slices of the same person. Although the re-identification question has been central to philosophical reflection about personal identity, Schechtman argues that there is another question about it. This second question is not about the metaphysical conditions of the persistence of persons through time, but rather about identity in a sense closer to the ordinary usage of the term ‘identity’. When in ordinary life people ask questions about their own and others’ identity, what is typically at issue is what they are like, what makes them distinctive as the persons they are. In contrast to identity as re-identification,

the question about identity in the sense of characterization inquires into the characteristics and psychological features that make a person the person she is (Schechtman, 1996, pp. 2, 74).

The re-identification and the characterization questions are logically different because they involve different relata. Whereas the former question is concerned with the relation between two person-slices, the latter question seeks to define the relation between a person and characteristics, actions, and experiences that are hers (Schechtman, 1996, p. 77). The difference between the two questions can be made vivid by considering that identity in the sense of characterization is the relevant notion of identity in the talk of an “identity crisis” (Schechtman, 1996, p. 74). When somebody has an identity crisis, what is typically at issue is not the kind of identity that depends on whether two person-stages are constituents of the same person, but rather identity understood as the attributability of certain actions, experiences, and psychological traits. More generally, whereas identity in the sense of re-identification is primarily about metaphysics, identity in the sense of characterization is about attributability and practical concerns. According to Schechtman, a narrow focus on the re-identification question is unable to capture some important pre-theoretical intuitions that link personal identity with issues pertaining to survival, moral responsibility, self-interested concern, and compensation (Schechtman, 1996, pp. 2, 149). Re-identification theorists who make implicit appeal to the cogency of those features are led astray by conflating the metaphysical question about identity in the sense of re-identification with the practically oriented question of identity in the sense of characterization. As a way out of this impasse, Schechtman proposes that giving centre stage to the characterization question is the right way to vindicate the practical significance that people ordinarily attach to their and other people’s identity.

For the purposes of this chapter, it is not central to consider how exactly to understand the relationship between the re-identification and the characterization questions. What is more relevant is that they are distinct questions that—going back to my guiding example—can also be raised about we-identities like the one Alma and Milton instantiate. It is one thing to ask how independently definable parts or slices of a group of friends are brought together into a numerically single entity that persists over time. It is another thing to ask which traits, characteristics, and actions are properly attributable to the group, or taking the insider’s perspective, to *us*. Suppose that, because of a serious fight which threatens the continuity of their friendship, Alma and Milton have to address the question of who they are as friends. There is a relevant distinction to be made between asking about the characteristics and features that make them who they are as friends, constituting them as a locus of practical significance, and asking about the conditions under which the group that they constitute is numerically identical with respect to, say, two months ago. The difference between the re-identification and the characterization questions in the context of we-identity can be made vivid by considering that—as it happens—identity crises come not only in the first-person

singular, but also in the first-person plural. A group of friends undergoing a relationship crisis may be very much concerned with the question ‘who are we?’, in a sense that has little to do with determining the re-identification criteria of a specific entity, and more with the characteristics that are rightly attributable to them.

The question about characterization is suitable to address one central feature of personal we-identities that I mentioned at the outset. I suggested there that personal we-identities are experienced from within. But what does it take for a we-identity to be so experienced? What does it take for me to experience myself as one of us? One requirement that has to be in place is that one’s membership is something that one subjectively appropriates, in the sense that one attributes it to oneself and incorporates it in one’s self-understanding. Importantly, attributability and appropriation play a central role in characterization issues, but not in re-identification issues. As Schechtman writes,

[a]n identity in the sense of the characterization question, is not [...] something that an individual has whether she knows it or not, but something that she has *because* she acknowledges her personhood and appropriates certain actions and experiences as her own.

(Schechtman, 1996, p. 95)

This aspect of acknowledgement and appropriation is arguably absent in the re-identification question. Whether or not a person is a numerically identical entity through time is a metaphysical issue that doesn’t depend on a subjective appropriation. Consequently, the characterization question is more suitable to examine the status of personal we-identities than the re-identification question.⁹ Pursuing further this line of reflection leads to consider how, if at all, there can be not only a subjective appropriation of events, values, etc., from a first-person singular perspective, but also intersubjective practices of appropriations from a first-person plural perspective.

Now, supposing that the characterization question is central for better understanding personal we-identities, what should the answer to that question be? Consider self-identity. According to Schechtman and other authors, the answer to the characterization question about self-identity is a *self-narrative* (Schechtman, 1996; see Kind, 2015, p. 130). There are significant disagreements in the literature about the very notion of narrative and the role of narratives in self-identity. Some take a self-narrative to be a precondition for selfhood, others hold that narratives are a widespread but not necessary tool for understanding oneself (and others). Some take narrative unity to be a *desideratum* that one should strive for, and others take narrativity to be a descriptive feature of how ordinary people experience their lives.

I don’t intend to address the many issues surrounding the notion of narrative here. Instead, I will adopt a modified version of the Narrative Self-Constitution View (Schechtman, 1996, 2011).¹⁰ The central idea of this view is that “we constitute ourselves as selves by understanding our lives as

narrative in form and living accordingly” (Schechtman, 2011, p. 398). The version I adopt is modified in two senses. First, I take narrative identity to be only one aspect or dimension of self-identity pertaining to personhood and self-understanding, and based on a more basic—yet less substantive—notion of minimal selfhood that pertains to phenomenal consciousness and self-experience (Zahavi, 2014). Second, it is not part of the notion of narrative that I adopt that there must be a single overarching narrative—a master narrative—spanning the whole of an individual’s person’s life. There might be more than one narrative, across different timescales, even though there might also be pressure towards narrative unification in a single life narrative.

A self-narrative, understood along the lines of the Narrative Self-Constitution View, is an unfolding developmental structure in the light of which specific events, episodes, and happenings gain their significance. More in detail, a self-narrative has a *structure*, a *holistic character*, and is partly *communicable*. First, to have a narrative self-identity is to conceive of some of the events and happenings of one’s life as articulated within a broader structure. Second, narratives are *holistic*, in the sense that they are structural wholes in terms of which individual parts gain their significance. Narratives are not made up of interconnections between independently definable parts, but are rather wholes in terms of which individual parts gain their intelligibility. As Schechtman illustrates, the experience of winning the lottery will be different for someone immensely wealthy, someone very poor, and someone who has unsuccessfully struggled with a gambling addiction (2011, p. 398). Finally, narratives have a communicative dimension. Even if some narratives may operate at an implicit level, a narrator must be in a position to articulate a narrative linguistically and communicate it, at least partially. The notion of a fully implicit self-narrative would be somewhat mysterious.

The just sketched conception of narrative has the advantage of avoiding two unattractive positions (see Schechtman, 2011). On the one hand, a notion of narrative that is too narrow and demanding. On the other hand, one that is too broad and permissive. If having a narrative identity would require from someone to constantly tell the events of her life in an actually written or told story, having a narrative identity would be a quite exceptional and rare accomplishment. On the other hand, if one opts for a broader and looser notion of narrative, the notion of narrative would become quite diluted and potentially trivial. If any actions that a subject performs would be *per se* narrative or part of a narrative, the narrative view appears to become vulnerable to the charge of “pan-narrativism” (Gallagher & Hutto, 2019, p. 33). As Gallagher and Hutto express this worry, this is “the idea that all things – the world in-itself, human lives – are, at root, narrative in nature” (2019, p. 33; see Strawson, 2004).

One way of responding to the charge of pan-narrativism is by conceptualizing narratives as representational artefacts produced for communicative functions (Gallagher & Hutto, 2019, p. 33). On this view, a narrative selects

and arranges contents in a certain way for communication. One might worry that this proposal relapses into the problem that narratives are exceptional accomplishments. But this need not be the case, because the requirement that narratives have a communicative function can be interpreted in at least two ways. On a strong reading, narratives have to be actually told in order to be operative. On a weaker reading, narratives may be largely implicit (in the sense of not being actually communicated), even if for a narrative to be operative particular stretches or segments of the narrative would have to be communicable (see Schechtman, 1996).¹¹

In sum, the concept of narrative tracks an important dimension of self-identity linked to characterization and attributability issues. It highlights that one central aspect of self-understanding doesn't come about by putting together independently definable pieces (episodes, events, traits, values, etc.), but rather by appropriating them within the temporally extended unity of an ongoing story. I contend that this conception of narrative can fulfil a double role. First, it can help elucidate personal we-identities exemplified by relations of companion friendship. Second, it can provide a common ground to understand the articulation between self-identity and we-identity. Since I will address the first point in some detail in the next section, I close this section by considering the second point.

Why would narrative identity be a suitable common ground to understand the articulation between self-identity and we-identity? As noted by several authors (Bruner, 1987, p. 21; MacIntyre, 1985; Ricoeur, 1990), one shouldn't think of narrative self-identity as a solitary accomplishment. Other people typically figure in important ways in self-narratives, not only as more or less anonymous sources of values, thoughts, and expectations, but also as more or less significant and concrete others: partners, parents, friends, colleagues, acquaintances, etc. This suggests that one reason why narrativity is a suitable common ground between self-identity and we-identity is that there is an important connection between narrative self-understanding and alterity. The narratives that *others* conceive and tell about oneself can have an impact on one's self-conception, and concrete others can also play a central role in one's individual self-conception. Going one step further, other people might also be important for narrative understanding because a plurality of individuals can also co-construct narratives together, as narratives from a first-person plural perspective.

4 We-narratives below corporate groups

I have motivated the idea that the characterization question about identity is suitable to clarify the status of we-identities, and I have introduced the notion of narrative as it plays out in discussions on self-identity. This section elaborates on the relationship between narrative understanding and we-identities, by focusing on the notion of we-narratives, introduced in the recent literature by Gallagher and Tollefsen. They describe we-narratives as "narratives

about what we are doing, have done, and will do, or what we ought to do and want to do” (Gallagher & Tollefsen, 2019; Tollefsen & Gallagher, 2017), and they motivate their account by considering the problems of the stability and the depth of shared agency (see Bratman, 2006). In brief, the problem of stability arises because a plurality of agents may have a prior shared intention to do something together even if they have different, and perhaps conflicting reasons to do it. This raises the concern of what can secure the stability of a shared intention. The problem of depth is that even if some agents might share a common framework of background reasons to do something together, they may each have different reasons for accepting that common framework.

Gallagher and Tollefsen propose that one way in which the stability and depth of shared agency can be increased is by means of *we*-narratives. More specifically, they suggest that *we*-narratives have three roles in the formation and maintenance of a *we*. The first role concerns the group’s reflective attribution of shared agency. When members of a group engage in retrospective and prospective narrative processes, they often do so by talking about what we have done or what we will do. Importantly, Gallagher and Tollefsen suggest that this is a way in which the group attributes agency to itself, given that a group narrative may extend “beyond any passing set of participants” (Gallagher & Tollefsen, 2019, p. 214). The *we*-narrative of a group secures that the latter is robust enough that the agential status of the group becomes independent of who the specific members of the group are at a given time. Consider, for example, a university board deciding on some modifications of curricula. What the board members decide together *qua* board members, deliberating from a first-person plural perspective and in terms of what *we* decide as the board members, is supposed to be independent from who the individual members of the board are at a given time.

A second role of *we*-narratives identified by Gallagher and Tollefsen concerns the group’s identity over time. *We*-narratives support the group’s identity because they can make it clear to others “who they are in terms of group identity”. This again may happen “even through a changing series of participants” (Gallagher & Tollefsen, 2019, p. 214). Consider again a university board, this time reflecting about the long-term mission and vision of the institution. The board can thereby create an “identity narrative” (Gallagher & Tollefsen, 2019) about the university. This may include a set of guiding principles and pedagogical objectives that make explicit what distinguishes the profile of that institution from other institutions. Finally, *we*-narratives about joint projects—particularly when these are long-termed planned—can also provide stability to *we*-formations. *We*-narratives help to explain why individuals can be bound together and continue their participation in a joint action, even if they have formed shared intentions for different reasons. This is because *we*-narratives can specify the content of commitment and assurances that motivate people to continue their participation in a joint action, and they may also indicate how they are going to proceed by specifying relevant means and goals (Gallagher & Tollefsen, 2019, p. 215).

Although Gallagher and Tollefsen suggest that the we relevant for their analysis can be “as small as a dyad and as large as a nation-state or an ethnic group” (Tollefsen & Gallagher, 2017, p. 102), it is fairly clear that their account of we-narratives is most readily applicable to *corporate groups*. In fact, they also write that

[a]ll three of these issues—joint agency, identity, and stability—are relevant to the concept of a *corporate* or instituted group, as distinguished from an *aggregate* (a collection of people not acting together) or a *plural* group (short-lived, constituted in one action—you and me moving a table together).

(Gallagher & Tollefsen, 2019, p. 215; see Gallagher, 2020, p. 221)

As mentioned in Section 2, corporate groups are taken to have a structure, a hierarchical organization, decision-making processes, and an identity that persists across changes of membership (Tollefsen, 2015; Tollefsen & Gallagher, 2017, p. 104). In contrast, an aggregative group is a collection of individuals with a common property, such as the collection of all red-haired women, and a plural group comes into existence in one single joint action (Tollefsen, 2015, p. 47).

Although Gallagher and Tollefsen don't claim that this taxonomy is exhaustive, one difficulty that arises is that it isn't clear how the group of friends constituted by Alma and Milton could fit into it. Their group is not a mere aggregate, because there doesn't seem to be one specific property that they have in common, and in virtue of which they would be members of their group. They are neither a plural agent, because their friendship is not a matter of one single joint action. And they are not a corporate group, because their individual membership in the group is essential. As pointed out above, the identity of a dyadic friendship like theirs does not persist through changes of membership.

Might we-narratives, nonetheless, play a role in the personal we-identity that Alma and Milton instantiate? I suggest that Gallagher and Tollefsen's notion of we-narratives is of relevance to answer this question, provided that some adjustments are made to the notion of we-narrative that they offer. First, while the reflective attribution of shared agency via retrospective and prospective narrative processes plays an important role in a friendship, it would not concern the group beyond any passing set of participants. Second, consider group identity. Whereas it is up to a university board to construct and reach agreement about a suitable identity narrative, it is much less clear that identity narratives are something that friends are typically concerned with. Friends may reflect jointly and explicitly about their 'we', but differently from corporations and institutions, that might happen if there is some conflict or uncertainty about who they are. At any rate, the idea of reaching agreement about who we are as friends seems problematic, insofar as it is not fully up to the friends to decide and draw

the boundaries of who they are. One can decide to cultivate a friendship and to do many things for one's friends, but a friendship as such cannot be established in the first place by acts of the will. It is relevant not to overlook the role of passivity and organic growth in the constitution of some affectively loaded social formations.¹² This point can be brought into focus by considering the distinction between the phenomena that Aron Gurwitsch calls *partnership* and *communal being-together*. Whereas the former is an instrumental association between people who voluntarily undertake a certain task together (and who are only bounded by it), the latter is not a matter of decision, will, and expected outcomes, but rather of growing together in a specific community (Gurwitsch, 1979).¹³

What about the issues of stability and depth of shared agency? I suggest that in the context of groups such as the one constituted by Alma and Milton, these problems are, to some extent at least, solved in advance. This is because what binds the friends in what they do together as friends is, at least partly, the flourishing of their relation. This is something that both friends value and want to promote. This, of course, is not to deny that we-narratives may play an important role indicating how friends—like any other co-agents—should proceed doing something together. The point is rather that concerns about the stability and depth of shared agency are at least considerably diminished once shared agency is placed in the context of concrete and normatively loaded interpersonal relations, such as friendship, which provide participants with reasons *for us* that are arguably absent in other social relations. In the next section, I discuss in more detail how we-narratives can provide an answer to the characterization question about personal we-identities exemplified by interpersonal friendship, and offer support for this proposal.

5 Historicity, communication, and the depth of friendship

My thesis is that we-narratives play a role in the constitution of a central form of friendship, like the one exemplified by Alma and Milton, because the unity of their social relation is best understood in terms of a *narrative* unity. This means that, as part of being friends, Alma and Milton conceive of some events, actions, and characteristics as articulated within an ongoing story. This story is not just a collection of items, but a broader unitary structure in the light of which particular events and characteristics are made sense of. On this view, the we-narrative of two companion friends does not consist of the specific episodes they have gone or will go through together, but rather in the way in which those episodes are appropriated and made intelligible in light of a more or less coherent and meaningful whole, which is the trajectory of their relation. In this section, I elaborate on three features of friendship—historicity, communication, and depth—and explain how the proposal that

the we-identities of friendship are narratively constituted helps to account for them.

One feature that makes friendships different from fleeting relations that arise and dissolve quickly is that the former are backward and forward-looking relations. Historicity matters for friendship because the shared history of a group of friends can be lived from within by them, individuating their group in a unique way and distinguishing them from other groups. Such shared history may be jointly organized, appropriated, and recalled in certain ways that are group specific. Research on transactive memory systems lends support to the idea that intimate social groups, like friendships, develop processes of memory encoding and retrieval that may be distributed across the group members, in such a way that no individual members can recall what they can jointly remember (Harris et al., 2011; Theiner, 2013). Moreover, the shared history of a friendship can be scaffolded by environmental features in specific ways. Far from being a self-enclosed mental construction, a friendship is a world-directed social formation. It can have its own specific “surrounding world,” constituted by aspects of the socio-material environment that are imbued with practical and affective significance for that particular social group (Husserl, 1952, pp. 195–196). Be it recurrently visited places or material objects with a special significance, environmental features can support the continuity of a friendship identity over time by providing members with opportunities for remembering a shared past and planning for the future together.

We-narratives can readily account for the historicity of friendship, because narrative identity is diachronically extended. It provides a framework within which the changes that an interpersonal relation undergoes across time can be made sense of, while preserving an overarching coherence. Moreover, we-narratives can also account for the role of environmental scaffoldings in friendship. Analogously to how, at the individual level, networks of objects that are open for narrative appropriation and ‘emplotment’ can support the continuity of a personal identity over time (Heersmink, 2018, p. 1836), there can also be group-specific networks of objects that are open for narrative appropriation and that support the identity of a group over time.

That being said, since the identity of other kinds of groups may be historically constituted and scaffolded by the environment in various ways, it is relevant to have a closer look at the case of friendship. One feature of friendship that sets it apart from prototypical corporate agents, like a university board, is that there is a sense in which the attribution of specific characteristics to the we-narrative of a friendship is geared towards the flourishing of the relation itself. This point can be brought into focus by considering some remarks by Stanley Cavell in the context of his analysis of Hollywood comedies of “re-marriage” from the 1930s. Cavell suggests that a characteristic feature of friendship is that friends care more about being together than about the

specific actions that they carry out. He calls this “the ascendancy of being together over doing something together” (Cavell, 1981, p. 88). As he writes in his reading of Frank Capra’s movie *It Happened One Night*:

What this pair does together is less important than the fact that they do whatever it is together, that they know how to spend time together, even that they would rather waste time together than do anything else—except that no time they are together could be wasted. Here is a reason that these relationships strike us as having the quality of friendship [...].
(Cavell, 1981, p. 88)

Implicit in Cavell’s remark that “no time they are together could be wasted” is the idea that being together as friends is something inherently valuable for the friends. For comparison, consider corporate groups. One reason why the university board might be concerned with the group identity of the institution is that the latter has a mission to carry out. In that sense, drawing the boundaries of the institutional we is relevant because of long-term institutional goals. What the mission of a companion friendship may be is considerably less clear. While one might want to make room for purely instrumental friendships (e.g. tennis buddies who meet regularly only because of the individual benefits they each extract from playing together), there is at least one central variety of friendship that is not intrinsically dependent on the achievement of specific goals external to the flourishing of the friendship itself.

This suggests that the kind of unity of friendship involves some recognition of the flourishing of the relevant we-identity as a good to be pursued for its own sake. In other words, there is an important sense in which friends are together for the sake of being together. Their community is not *necessarily* structured around external objectives or goals—although it might be derivatively so structured—but is geared towards the flourishing of their communal relation. As distinguished from “objectual communities”, friendship can be qualified as a “reflexive community” in which “[t]he community and the life in the community become [...] an end in itself [*Selbstzweck*]” (Walther, 1923, p. 67. My translation; see also Löwith, 2016/1928, pp. 147, 160). In-consequential and trivial endeavours, and the mere ‘hanging around’ with each other can be valued as such by the friends because those episodes are interpreted as parts of the ongoing we-narrative of their friendship, to which they attach some intrinsic value.

Both common sense and philosophical reflection suggest that friends stand in a communicative relation to one another and that lack of communication erodes friendships (Aristotle, 2019 – 1157b13). I think we should uphold these two ideas. Although there surely is a lot of variation in patterns of communication between friends, the idea of being friends with someone with whom the possibility of reciprocal communication is foreclosed doesn’t make much sense. But assuming that the adoption of a communicative stance towards

each other is a constitutive aspect of a relation of friendship, what account of the we-identity of friendship could illuminate it? Given that narratives have a communicative dimension, the narrative approach is in a position to do so. Communicating subjects can jointly build a conception under which they value their being together, co-constructing and extending their common ground, and conceiving of themselves as *us* by locating specific characteristics in the context of their shared history. This, I suggest, is not different from co-constructing the we-narrative of their relationship. At the same time—like self-narratives—we-narratives need not be constantly told to be operative. But they have to be partly *communicable*. Such we-narratives typically involve retrospective recollection and prospective deliberation about what we have done or will do, but they may also operate implicitly.

Finally, consider the *depth* of friendship. By this I mean that being ‘one of us’ as friends is not just like any other property that an individual happens to instantiate, but rather one that potentially shapes who she or he is in a profound way. The lives of close friends interpenetrate in complex ways, which go beyond a self-categorization dependent on contextual cues (Turner et al., 1994), and also beyond the self-disclosure of confidential information—something that may happen reciprocally between, say, two psychoanalysts who analyse each other (Reiman, 1976, p. 33). Friends are prone to see each other through the other’s eyes, and they tend to develop “relational identities” (Cocking & Kennett, 1998; Cocking & Matthews, 2000), insofar as they shape aspects of their individual identities in and through their interpersonal relation. This becomes particularly evident in disruptions and dissolutions of friendship. As Nehamas suggests, the grief of being deserted by a friend is peculiar in that it isn’t just about what one friend might have done. Rather, “[a]t the core of such grief lurks the anguish provoked by a deeper, more pervasive abandonment, a rejection not merely of what we have *done* but of what we *are*” (Nehamas, 2010, p. 286).

How to account for the tight relation between friendship and self-identity? One possibility is to argue that the depth of friendship has a distinctive normative character. Consider the proposal that close friends come to constitute a “plural agent” (Helm, 2008, 2010), insofar as they come to jointly hold a shared evaluative perspective. Even though plural agents are said to be ontologically dependent on their members (Helm, 2010, p. 265), they are supposed to have specific patterns of rationality and to be agents that exist in their own right. In a plural agent, the mental states of some individuals are rationally interconnected in such a way that they are responsive not to what matters to them qua individuals, but rather qua *one of us* (Helm, 2010, p. 276). It is the group as a whole that is the subject of import, and each member of the plural agent can be held rationally accountable for his responses in the light of the joint evaluative perspective (2010, p. 273). Participation in such a plural agent could potentially explain that being ‘one of us’ as friends is a property with repercussions on self-identity.

But one shortcoming of this proposal is that it leaves rather unexplained how a plural agent comes about. How does a plural agent relate to the

attitudes of the individual agents that constitute it? To put it differently, what comes first: the plural agent or the attitudes towards the plural agent? Pointing out that there is a circularity here, but not of a vicious kind because “each comes into existence simultaneously” (Helm, 2010, pp. 281–282), is not a satisfactory solution. In fact, this emphasis on simultaneity doesn’t square well with the historicity of some putative plural agents, including those exemplified by relations of companion friendship. The latter are backward- and forward-looking relations that are diachronically extended and build on a shared history.

I suggest that the notion of we-narratives can be of help here. Although the depth of friendship may well have a normative character, it is one that arises historically, from the progressive interpenetration of self-narratives under an emerging we-narrative. As proposed above, the latter depends on the communicative appropriation of certain characteristics as ours, which is a process of narrative construction. The narrative approach offers resources to explain the interpenetration of friends’ self-identities in terms of the interrelation between self-narratives that become integrated within an overarching we-narrative that develops over time. This proposal can explain why relations of friendship can not only greatly enrich self-identity, but also—and more interestingly—partly constitute it. For example, co-constructing a we-narrative with Milton, and partaking in a we-identity of friendship with him, can be a characteristic with which Alma comes to strongly identify *qua* individual person. Indeed, it can become a constitutive feature of her self-identity and her self-narrative, to the extent that losing that feature can be experienced as a loss of one part of her identity.

6 Concluding remarks

By way of conclusion, I come back to my initial question: what is the relationship between the I and the we? One aim of this chapter has been to show that the traditional focus on collective intentionality can be fruitfully supplemented with a focus on we-identity. A second aim has been to show how this can be done by investigating the personal we-identities characteristically present in one central form of friendship. I have argued that the narrative approach offers resources to account for the we-identities of companion friendship, insofar as it helps to better understand the historicity, communicative character, and depth of the latter. Whether this proposal can also be applied to other types of interpersonal relations is a question that remains open for further exploration.

Acknowledgements

Thanks to Dan Zahavi, Mikko Salmela, and Mette Vesterager for comments on earlier versions of this chapter. The author acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 832940).

Notes

- 1 Center for Subjectivity Research, Department of Communication, University of Copenhagen, Karen Blixens Plads 8, DK-2300 Copenhagen, Denmark. E-mail: felipe.leon@hum.ku.dk
- 2 On the Aristotelian motif, widely accepted in contemporary discussions on friendship, that friends care for each other for the other's own sake, see (Aristotle, 2018 – 1380b35–1381a, 2019 – 1156b7–12; Brink, 1999). Aristotle distinguishes between three kinds of friendship—friendships of *utility*, of *pleasure*, and of *character*—and attributes the feature of caring for the other for the other's own sake to the last kind, which he takes to be the highest and most complete kind of friendship. For the purposes of this chapter, I take the present example to illustrate one central variety of friendship, and I leave open that there might be other varieties of it (for discussion, see Helm, 2017).
- 3 Although I believe that these features can also be found in other types of interpersonal relations, including relations of romantic love, I will focus on companion friendship.
- 4 The proposed taxonomy is not intended to exclude cases in which the different categories of we-identity may partly overlap.
- 5 For arguments in support of the claim that core cases of cooperative *telling* are joint actions, based on the normative properties of telling and on deception counterexamples to Gricean accounts of communication, see (Jankovic, 2018, 2014).
- 6 Some theorists highlight the persistence of the group's identity across changes of membership because they take it to be a precondition for the attribution of rights and responsibility to the group as such. Another issue that I don't consider here is that some corporate groups may require for their persistence the stable membership of *specific* individuals.
- 7 Let me note here that a further variety of we-identity might be linked to ethnic and national groups, which persist across changes of individual membership, but lack any clear decision-making procedures, hierarchical organization, and other features typically associated to corporate groups.
- 8 Needless to say, the label 'personal we-identities' is not intended to suggest that ephemeral integrations and corporate we-identities don't involve persons and interactions between them. That would be obviously false. The point is rather that neither ephemeral nor corporate we-identities have as a necessary precondition that individuals relate to one another *personally*.
- 9 For an account of collective identity that combines aspects of both the characterization and the re-identification questions, see (Mathiesen, 2003).
- 10 For recent elaborations and discussion of this view, see (Heersmink, 2018; Walker, 2019).
- 11 Another way to avoid the problem of exceptionalism is by taking narratives to be heuristic tools that we occasionally and locally employ to make sense of ourselves (and others), and that might play a critical role in the acquisition of socio-cognitive competences (Hutto, 2016). Although I think there is a lot to be said in favour of this proposal, one downside of it is that it de-emphasizes the link between narrative understanding and self-identity, which is more important for the purposes of this chapter.
- 12 For an interesting discussion of the limits of voluntary self-disclosure in the context of 'online' friendships, see (Cocking & Matthews, 2000).
- 13 Although Gurwitsch doesn't discuss companion friendship, his distinction bears on the present point. Gurwitsch draws here on Tönnies' influential distinction between community (*Gemeinschaft*) and society (*Gesellschaft*) (Tönnies, 2019/1887).

Bibliography

- Aristotle. (2018). *Rhetoric* (C. D. C. Reeve, Trans.). Indianapolis, IN: Hackett Publishing Company.
- Aristotle. (2019). *Nicomachean Ethics* (T. Irwin, Trans.). Indianapolis, IN: Hackett Publishing Company.
- Bratman, M. (2006). Dynamics of Sociality. *Midwest Studies in Philosophy*, 30(1), 1–15. <https://doi.org/10.1111/j.1475-4975.2006.00125.x>.
- Brink, D. O. (1999). Eudaimonism, Love and Friendship, and Political Community. *Social Philosophy and Policy*, 16(1), 252–289. <https://doi.org/10.1017/S0265052500002326>.
- Bruner, J. (1987). Life as narrative. *Social Research*, 54(1), 11–32.
- Calabi, C. (2008). Winks, Sighs and Smiles? Joint Attention, Common Knowledge and Ephemeral Groups. In H.-B. Schmid, K. Schulte-Ostermann & N. Psarros (Eds.), *Concepts of Sharedness: Essays on Collective Intentionality*. Heusenstamm: Ontos Verlag, pp. 41–58.
- Carr, D. (1986). *Time, Narrative, and History*. Bloomington: Indiana University Press.
- Cavell, S. (1981). *Pursuits of Happiness: The Hollywood Comedy of Remarriage*. Cambridge, MA: Harvard University Press.
- Cocking, D., & Kennett, J. (1998). Friendship and the Self. *Ethics*, 108(3), 502–527. <https://doi.org/10.1086/233824>.
- Cocking, D., & Kennett, J. (2000). Friendship and Moral Danger. *The Journal of Philosophy*, 97(5), 278–296. <https://doi.org/10.2307/2678396>.
- Cocking, D., & Matthews, S. (2000). Unreal Friends. *Ethics and Information Technology*, 2(4), 223–231. <https://doi.org/10.1023/A:1011414704851>.
- Epstein, B. (2019). What are Social Groups? Their Metaphysics and How to Classify Them. *Synthese*, 196(12), 4899–4932. <https://doi.org/10.1007/s11229-017-1387-y>.
- Frankfurt, H. (1988). *The Importance of What We Care About: Philosophical Essays*. New York: Cambridge University Press.
- French, P. A. (1984). *Collective and Corporate Responsibility*. New York: Columbia University Press.
- Gallagher, S. (2020). *Action and Interaction*. New York: Oxford University Press.
- Gallagher, S., & Hutto, str. (2019). Narratives in Embodied Therapeutic Practice. Getting the Story Straight. In H. Payne, S. C. Koch, J. Tantia, & T. Fuchs (Eds.), *The Routledge International Handbook of Embodied Perspectives in Psychotherapy: Approaches from Dance Movement and Body Psychotherapies*. New York: Routledge, pp. 28–39. <https://doi.org/10.4324/9781315159416>.
- Gallagher, S., & Tollefsen, D. (2019). Advancing the ‘We’ Through Narrative. *Topoi*, 38(1), 211–219. <https://doi.org/10.1007/s11245-017-9452-1>.
- Gurwitsch, A. (1979). *Human Encounters in the Social World*. Pittsburgh: Duquesne University Press.
- Harris, C. B., Keil, P. G., Sutton, J., Barnier, A. J., & McIlwain, D. F. J. (2011). We Remember, We Forget: Collaborative Remembering in Older Couples. *Discourse Processes*, 48(4), 267–303. <https://doi.org/10.1080/0163853X.2010.541854>.
- Heersmink, R. (2018). The Narrative Self, Distributed Memory, and Evocative Objects. *Philosophical Studies*, 175(8), 1829–1849. <https://doi.org/10.1007/s11098-017-0935-0>.
- Helm, B. W. (2008). Plural Agents. *Nous*, 42(1), 17–49. <https://doi.org/10.1111/j.1468-0068.2007.00672.x>.

- Helm, B. W. (2010). *Love, Friendship, and the Self: Intimacy, Identity, and the Social Nature of Persons*. New York: Oxford University Press.
- Helm, B. W. (2017). Friendship. In *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/friendship/>.
- Husserl, E. (1952). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Zweites Buch. Phänomenologische Untersuchungen zur Konstitution*. Dordrecht: Martinus Nijhoff.
- Hutto, D. (2016). Narrative Self-Shaping: A Modest Proposal. *Phenomenology and the Cognitive Sciences*, 15(1), 21–41. <https://doi.org/10.1007/s11097-014-9352-4>.
- Jankovic, M. (2014). Communication and Shared Information. *Philosophical Studies*, 169(3), 489–508. <https://doi.org/10.1007/s11098-013-0205-8>.
- Jankovic, M. (2018). Telling and Mutual Obligations in Communicative Action. *ProtoSociology*, 35, 99–114. <https://doi.org/10.5840/protosociology2018356>.
- Kind, A. (2015). *Persons and Personal Identity*. Cambridge: Polity Press.
- Konzelman-Ziv, A. (2009). The Semantics of Shared Emotion. *Universitas Philosophica*, 26(52), 81–106.
- Krueger, J. (2016). The Affective “We”: Self-regulation and Shared Emotions. In T. Szanto & D. Moran (Eds.), *Phenomenology of Sociality: Discovering the We*. New York: Routledge, pp. 263–277.
- León, F., Szanto, T., & Zahavi, D. (2019). Emotional Sharing and the Extended Mind. *Synthese*, 196, 4847–4867. <https://doi.org/10.1007/s11229-017-1351-x>.
- List, C., & Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. New York: Oxford University Press.
- Löwith, K. (2016/1928). *Das Individuum in der Rolle des Mitmenschen*. Freiburg/München: Verlag Karl Alber.
- MacIntyre, A. (2007). *After Virtue: A Study in Moral Theory*. 3rd Edition. Notre Dame, IN: University of Notre Dame Press.
- Mackenzie, C., & Atkins, K. (Eds.). (2008). *Practical Identity and Narrative Agency*. New York: Routledge.
- MacIntyre, A. (1985). *After Virtue. A Study in Moral Theory*. London: Duckworth.
- Mathiesen, K. (2003). On Collective Identity. *ProtoSociology*, 18, 66–86. <https://doi.org/10.5840/protosociology200318/192>.
- Mathiesen, K. (2005). Collective Consciousness. In D. W. Smith & A. L. Thomasson (Eds.), *Phenomenology and Philosophy of Mind*. New York: Oxford University Press, pp. 235–250.
- Millgram, E. (1987). Aristotle on Making Other Selves. *Canadian Journal of Philosophy*, 17(2), 361–376. <https://doi.org/10.1080/00455091.1987.10716441>.
- Nehamas, A. (2010). The Good of Friendship. *Proceedings of the Aristotelian Society*, 110(3pt3), 267–294. <https://doi.org/10.1111/j.1467-9264.2010.00287.x>.
- Pacherie, E. (2011). The Phenomenology of Joint Action: Self-Agency vs. Joint-Agency. In A. Seemann (Ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*. Cambridge, MA: MIT Press, pp. 343–389.
- Pacherie, E. (2014). How Does It Feel to Act Together? *Phenomenology and the Cognitive Sciences*, 13(1), 25–46. <https://doi.org/10.1007/s11097-013-9329-8>.
- Pettit, P. (2014). How to Tell if a Group Is an Agent. In J. Lackey (Ed.), *Essays in Collective Epistemology*. New York: Oxford University Press, pp. 97–121.
- Reiman, J. H. (1976). Privacy, Intimacy, and Personhood. *Philosophy and Public Affairs*, 6(1), 26–44. <https://doi.org/10.4324/9781315246024-3>.
- Ricoeur, P. (1990). *Soi-même comme un autre*. Paris: Seuil.

- Salmela, M., & Nagatsu, M. (2017). How Does It Really Feel to Act Together? Shared Emotions and the Phenomenology of We-Agency. *Phenomenology and the Cognitive Sciences*. <https://doi.org/10.1007/s11097-016-9465-z>.
- Schechtman, M. (1996). *The Constitution of Selves*. Ithaca, NY: Cornell University Press.
- Schechtman, M. (2011). The Narrative Self. In S. Gallagher (Ed.), *The Oxford Handbook of the Self*. New York: Oxford University Press, pp. 394–418.
- Scheler, M. (2008). *The Nature of Sympathy*. New York: Transaction Publishers.
- Schmid, H. B. (2009). *Plural Action* (Vol. 58). Dordrecht: Springer Netherlands. <http://link.springer.com/10.1007/978-90-481-2437-4>.
- Schmid, H. B. (2014). Plural Self-Awareness. *Phenomenology and the Cognitive Sciences*, 13(1), 7–24. <https://doi.org/10.1007/s11097-013-9317-z>.
- Schweikard, D., & Schmid, H. B. (2013). Collective Intentionality. In *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/collective-intentionality/>.
- Searle, J. R. (2010). *Making the Social World: The Structure of Human Civilization*. New York: Oxford University Press.
- Strawson, G. (2004). Against Narrativity. *Ratio*, 17(4), 428–452. <https://doi.org/10.1111/j.1467-9329.2004.00264.x>.
- Taipale, J. (2019). The Structure of Group Identification. *Topoi*, 38(1), 229–237. <https://doi.org/10.1007/s11245-017-9463-y>.
- Theiner, G. (2013). Transactive Memory Systems: A Mechanistic Analysis of Emergent Group Memory. *Review of Philosophy and Psychology*, 4(1), 65–89. <https://doi.org/10.1007/s13164-012-0128-x>.
- Tollefsen, D. (2004). Collective Intentionality. In J. Fieser & B. Dowden (Eds.), *Internet Encyclopedia of Philosophy*. <https://iep.utm.edu/coll-int/>.
- Tollefsen, D. (2014). A Dynamic Theory of Shared Intention and the Phenomenology of Joint Action. In S. R. Chant, F. Hindriks, & G. Preyer (Eds.), *From Individual to Collective Intentionality*. New York: Oxford University Press, pp. 12–33.
- Tollefsen, D. (2015). *Groups as Agents*. Cambridge: Polity Press.
- Tollefsen, D., & Gallagher, S. (2017). We-Narratives and the Stability and Depth of Shared Agency. *Philosophy of the Social Sciences*, 47(2), 95–110. <https://doi.org/10.1177/0048393116672831>.
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge: MIT Press.
- Tönnies, F. (2019/1887). *Gemeinschaft und Gesellschaft*. Berlin and Boston, MA: De Gruyter.
- Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and Collective: Cognition and Social Context. *Personality and Social Psychology Bulletin*, 20(5), 454–463. <https://doi.org/10.1177/0146167294205002>.
- Walker, M. J. (2019). Two Senses of Narrative Unification. *Philosophical Explorations*, 22(1), 78–93. <https://doi.org/10.1080/13869795.2018.1474943>.
- Walther, G. (1923). Zur Ontologie der sozialen Gemeinschaften. In E. Husserl (Ed.), *Jahrbuch für Philosophie und phänomenologische Forschung*, 6, 1–158.
- Wollheim, R. (1984). *The Thread of Life*. Cambridge, MA: Harvard University Press.
- Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy, and Shame*. New York: Oxford University Press.
- Zahavi, D. (2019). Second-Person Engagement, Self-Alienation, and Group-Identification. *Topoi*, 38(1), 251–260. <https://doi.org/10.1007/s11245-016-9444-6>.

12 A person's life

Uniting metaphysics and value

Jörg Noller

1 Human or person?

The question of what we actually are is controversially discussed in the current philosophical debate.¹ There are two possible answers: We are either basically *human animals* or *persons*. The first position – animalism – has biology on its side and is not problematic from an ontological point of view, since it presupposes nothing that could not be explained scientifically. Its disadvantage, however, is that as beings understood purely biologically, we can hardly claim a special position in the natural kingdom, let alone something like an inviolable dignity that belongs only to us as persons, not to *other* animals. However, we have strong intuitions for this special position, which cannot be given up simply out of biological considerations. It is exactly the other way around with the second position – which I will call “personalism” for the sake of simplicity. It argues that we are qualitatively distinguished as persons compared to other living beings – *mere* animals. Its drawback, however, is that this special status is difficult to assert against biology without making strong additional ontological assumptions. It seems, then, that we are not quite able to bring our being human and our being a person to a unified concept. The problematic relation between being human and being a person has implications for the question of our *identity*:

In everyday life we use the word ‘person’ in many different ways. Sometimes it means ‘human animal,’ sometimes ‘moral agent,’ sometimes ‘rational, self-conscious subject,’ sometimes ‘possessor of particular rights,’ sometimes ‘being with a defined personality or character,’ and there are many other senses as well. Each of these conceptions of person has its own corresponding criterion of personal identity, and there is no reason to assume that we can find some single relation which underlies our judgments about the identity of a ‘person’ in every context.²

Schechtman doubts that we can determine the concept of a person unambiguously and uniformly, and that such a uniform concept is based on a single relation that makes the identity of the person intelligible and justifiable.

Eric T. Olson has therefore argued that “[t]here is no single problem of personal identity, but rather a wide range of questions that are at best loosely connected”.³

In what follows, I will argue for a unified concept of the person, according to which not a single relation but a combination of relations establish the identity of the person. The aim of this chapter is to find a middle way between animalism and personalism and to develop a concept of a person’s life which avoids their problems.⁴ I will argue for the following thesis: As humans we are persons, but we are persons not in the sense of certain characteristics or abilities that distinguish our biological kind, but by a certain form of our life, without this form of life having to be limited only to us humans. Therefore, I will strictly distinguish between the biological *species* or *species-form* of human beings and the *life form* of humans as persons, which are often identified in the current debate.⁵

2 Animalism and constitutionalism

As described in the previous chapter, the question of “what we fundamentally are”, that is, what constitutes our identity, is controversial in the current debate. Although this debate knows quite different positions, it is particularly characterized by two opposing points of view – animalism and constitutionalism. Representatives of animalism, such as Eric T. Olson, argue that we are *essentially* (human) animals and only accidentally persons.⁶ Our conditions of identity or persistence are those of an animal organism. We exist only as long as this organism is alive, and with its death our existence ends. It follows that our identity conditions are not restricted to mental properties such as self-consciousness, which Locke, for example, had claimed. We may exhibit personal abilities at times, but viewed as a whole, especially with regard to the beginning and end of our lives, we are living organisms – nothing more, but also nothing less. I will call this thesis the “indifference thesis” of animalism.

Animalism is opposed by constitutionalism, whose main representative is Lynne Rudder Baker. The main intuition of constitutionalism is that it judges mental and moral properties to be so ontologically significant that a difference from mere organic life follows from it, which entails new conditions of identity. Constitutionalism argues that we as persons are basically not animals, as animalism claims. In what follows I will call this thesis the “difference thesis” of constitutionalism. According to the difference thesis, although psychological continuity is not sufficient for personal identity, it is the unity with our animal organism – a unity that does not imply identity – through which we are constituted as persons. The capacities we develop through gaining self-consciousness make us more than mere animals. Personhood is therefore not a mere attribute that may or may not belong to the animal organism, as animalism argues, but a property that constitutes an ontology distinct from it.⁷ Constitutionalism states that we are first and foremost persons, and only secondarily animals. Crucial for the status of a person is a specific mental qualification, which Baker determines as the capacity of a first-person perspective. As soon as an organism develops a first-person perspective,

a new entity appears in the form of the person, which is not identical with it.⁸ Accordingly, it is not biology that determines our conditions of identity, as is the case in animalism, but a specific ontology of subjectivity. Baker sums up her position by the formula "Ontology need not follow biology"⁹:

[B]iology does not fully reveal our nature. So, perhaps we should say that biology may well reveal our animal nature, but that our animal nature does not exhaust our nature all things considered. Rather, self-consciousness distinguishes us ontologically from the rest of the animal kingdom. This is to say that self-consciousness – and thus personhood – is an ontologically significant property.¹⁰

According to constitutionalism, this new entity of the person, which is constituted by but not identical with its organism, is not an entity existing in addition to the organism, as is the case with substance dualism.¹¹ Baker wants to go a middle path between animalism and substance dualism: The body organism and the person constituted by it have a first-person perspective, but in terms of a different *mode of having*. Baker also calls this modalism "key distinction".¹² According to it, the person has the first-person perspective in an underived, privileged way, while the body constituting the person has it only in a derived way, by "second-hand" as it were.

However, both constitutionalism and animalism have serious problems in determining exactly what we are. Constitutionalism, for instance, is affected by the "too many minds objection"¹³ or the "too many thinkers problem".¹⁴ This objection can be further plausibilized by the following "master argument"¹⁵ of animalism, on which the validity of animalism depends:

- (P1) Presently sitting in your chair is a human animal.
- (P2) The human animal sitting in your chair is thinking.
- (P3) You are the thinking being sitting in your chair.
- (C) Therefore, the human animal sitting in your chair is you.¹⁶

If we admit that there is an animal which thinks and which occupies the same place as we do, then we need to identify ourselves with this thinking animal. For otherwise there would be two thinking beings at the same place: A thinking animal and we – as a thinking person that is different from it. Hence, if we want to reject animalism, we have to reject one of the premises of the above argument with good reasons. One could now object here – as Baker does – that the animal, that is, our bodily organism, does not actually think, but thinks only in a derivative way, that is to attack P2. Following Baker, one could also attack P3 and distinguish between the person and (human) animal. However, Baker's modal distinction between the derived and the underived having of mental properties can counter the master argument only conceptually, but not ontologically. For how is a bodily organism supposed to have a first-person perspective in a derived way ontologically? Either it possesses it, or it does not possess it. If it possesses it, then we are thinking animals – nothing more,

but also nothing less. If not, then we seem to have to understand ourselves as thinking substances existing independently and separately from their bodily organisms, which presupposes a denial of P3. A “unity without identity”¹⁷ – the spatiotemporal coincidence of organism and person *without* their identity as advocated by Baker – is ontologically incomprehensible.

But animalism is not unproblematic either. For the assumption that we are essentially animal organisms conflicts with the practical self-understanding of our normative identity, according to which we as human beings have the capacity for free self-determination and thus enjoy an individual, inviolable dignity. Lynne Baker has formulated this problem of the practical and normative indifference of animalism as follows:

[A]ccording to Animalism [...], there is no ontological distinction between us and earthworms. By contrast, I think that metaphysics should tell us about what is fundamental to our being the kind of thing that we are (as opposed to earthworms), and about what is significant about us.¹⁸

Baker further criticizes Olson’s strict separation between metaphysics and value:

Olson wants to keep what is distinctive about us and what is most significant about us out of metaphysics. He does not discuss what is distinctive about us at all, and he consigns what is significant about us to a sphere of practical concerns outside the purview of metaphysics altogether.¹⁹

David Shoemaker has brought this problem of animalism, which consists in its practical underdetermination that follows from its indifference thesis, into the following argument:

- 1 Animalism lacks the proper fit with the set of our practical concerns;
- 2 If a theory of personal identity lacks the proper fit with the set of our practical concerns, it suffers a loss in plausibility; thus,
- 3 Animalism suffers a loss in plausibility (in particular to psychological criteria of identity).²⁰

Thus, while constitutionalism accommodates our practical self-understanding, insofar as it identifies the role of self-consciousness as ontologically significant for our identity but gets itself into serious ontological problems (such as the “too many minds objection”), animalism can only gain conceptual cogency by fully conforming to biology. As such, animalism ignores the practical and normative dimensions of our existence, which, however, are precisely co-decisive for our self-understanding – our individuality, autonomy, and dignity. Marya Schechtman has therefore rightly argued that the animalistic conception of life – as an organism – is only an abstraction of the life that persons lead.²¹ However, Schechtman does not specify further how exactly this complex life, which persons lead, must be thought ontologically. How, then, can we develop a concept of our personal existence

and persistence that is plausible both ontologically and normatively, and that avoids both the problems of animalism and constitutionalism? Eric T. Olson has pointed out the need for such a unified conception of the person:

[N]o account of our identity has yet been proposed that guarantees – in all possible cases, including science fiction – the coincidence of what is important in our identity with the actual conditions of our identity. [...] [A]n account of our identity must be ontologically coherent as well as ethically plausible.²²

In what follows, I will argue for a third way beyond animalism and constitutionalism concerning the identity of persons. In doing so, I will develop a notion of personal identity that avoids the problems of both theories and mediates between animalism and constitutionalism in ontological and practical regards. I will focus on the specific form of life that persons instantiate, realize, and share.²³ The alternative between animalism and self-conscious person is thus false. It is not capacities and properties that ontologically elevate mere animals to persons, but it is the morphology of their lives.

3 Species and life form

The identification of species and life form is already problematic from the point of view of biology, since the species describes a genealogical relationship, while the life form is a typological term,²⁴ i.e., it describes the structure which living beings exemplify across species by “their similarity in (external) shape and way of life”.²⁵ Thus, a life form is not about the contingent, diachronically grown, taxonomic relations of descent, but about “relations of form”²⁶ of the life of natural species that are to be analyzed synchronously. Life forms do not belong to biology alone, since they exceed it as ecological and, in the last consequence, social phenomena, even if they always remain related to it. The concept of life form is thus a thoroughly relational concept, whereby these relations can be further explicated in different ways. Ecologically, life forms, unlike species forms, are essentially context-relative. Their form gains profile in relation to their habitat, their lifeworld, and the role the species plays in it.²⁷ In this respect, one can speak of the life form as being a second-level form compared to the species form. For life forms always imply reflections and reactions to an environment, which leads to an adaptation and establishment in it.

Since the life form concerns relations of form which are not limited to the natural specifications of biology, it is a second-order form. This opens a transition to a genuinely philosophical consideration, which is nevertheless compatible with a biological description. The life form describes a structure which can be transferred to other areas of reality, such as persons, in such a way that at the same time the reference to their biological nature is given without being bound to its material specifications. The transition from species to life form is thus not a biological but a morphological one. In this

respect, two moments can be philosophically derived from the concept of life and made fruitful for the concept of the person: (i) The *biological* species concept or species form and (ii) the *biomorphic* life form concept.²⁸ The relation between these two forms raises the following three philosophical questions:

- 1 What is the ontological relation between the material particles and the individual (human) life under the species form?²⁹
- 2 What is the morphological relationship between species form and life form?
- 3 What is the normative relationship between the human species form and the personal life form?

It follows from the distinction between species and life form that the person is not *a* species concept, i.e., no species-specific properties or capabilities constitute the person. Therefore, one would look at the wrong place, if one already took the biological species as the person.³⁰

4 A person's form

The personal life form describes the mode in which quite different properties and capacities of an individual life are realized personally.³¹ What exactly it means that species-specific capacities are realized personally shall be exemplified by reference to Kant's theory of rationality. The distinction between instrumental (hypothetical) and moral (categorical) reason can be understood in such a way that the former concerns only the individual and its species-specific interests, which serve, for example, its own survival, while the latter denotes a more reflective form of rationality that involves other persons – by including and generalizing the interests of rational agents. In this respect, Kant determines a human in two ways: As a merely “rational *natural being*”,³² which stands under the form of hypothetical imperatives, but also as “a being capable of obligation [...] (humanity in its person)”,³³ which stands under the form of the moral law.³⁴ The faculty of rationality can thus be realized both naturally and personally. It is, morphologically speaking, both times the same faculty, but differently configured. This different use of the faculties shows that the personal life form does not ontologically separate the species form, but is always reflexively related to it.

Against this background, Kant's distinction between empirical and intelligible character can also be interpreted morphologically. The empirical character can be related to the specific species form “human”, which concerns the person as an individual natural being. The intelligible character, on the other hand, can be related to the general personal life form, because according to Kant the validity of the moral law does not only concern humans, but “all *rational beings in general*”.³⁵ The intelligible character, understood as the morphological structure of the personal life form, can be integrated or *infigured* into the empirical character – understood as the species form – in such a way that it does not describe a reality dualistically different from it, but rather determines the life under the species form “human” more precisely. Following the distinction

between species form and life form, the normativity of human rationality can be explained by the fact that human life is reflexively extended by the personal life form, which always presupposes relations and contexts – in Kant's case, for instance, the self-reflexive structure of the universalization of maxims. In this respect, species form and life form always remain related to each other, even if they do not immediately logically follow from each other.³⁶ This has the systematic advantage that through the personal life form the relation to the individual life of the species form as well as to a general structure encompassing the individual is given, which can be further explicated in a normative sense.³⁷

Morphologically speaking, the relation of species and life forms can be determined in such a way that the life form is a configuration of the individual life standing under the species form – thus a formal distinction of only certain species-specific traits, which lets the species stand in a non-genealogical-taxological relationship to other life forms – as, for example, the personal life form. This means that there exist other life forms than the personal life form, which can be realized across species, such as “predator” or “prey animal”.³⁸ This example shows that the life form starts from the species form, but is always relationally constituted. The relations of form that constitute it do not concern certain properties and capacities that are exclusive to the respective species, but their relation to surrounding contexts and other forms of life. Specifically, a predator is a predator only insofar as it is oriented toward the life form of the prey animal, just as the life form of the prey animal must always be understood in relation to that of the predator.³⁹ However, a special form of intersubjectivity and contextuality is decisive for the *personal* life form, which in the current debate is known under phrases like “person-space”⁴⁰ or “personal space”⁴¹ or “second nature”⁴². Therefore, on the one hand, it is necessary to develop this intersubjective structure of the personal life form in such a way that, on the one hand, through the morphological difference to the species form, a speciesism⁴³ and naturalistic fallacy is avoided, which already recognizes certain (moral) values in the nature (teleology) of the species (form).⁴⁴ On the other hand, however, the reference must be so arranged that still the reference to the identity of the species form is still kept, in order to avoid a dualism of first and second nature and the falling apart of organism and person. This can be achieved by understanding the life form as a reflection on the individual life under the species form, which configures it.

Seen in this way, the relationship between species and life forms does not imply a dualism of two forms, but an integrative combination of forms. The following scheme exemplifies the relation of species form and life form using the examples of “prey animal”, “predator”, and “person”:

Life form	Prey	Predator	Person
Species form	Deer, mouse, fly	Lion, owl, wasp	Human, ???
Characteristics	Escape and defense strategies and behavior	Attack strategies and hunting behavior	Normatively and reflexively-rationally qualified intersubjectivity

To illustrate the distinction between species form and life form, let us first consider the life form of the predator. It can be exemplified by quite different natural species and genera – mammals, birds, and insects. The life form abstracts from species-specific peculiarities of the species form, but at the same time shows in it a new and stable context of form. The life form of the predator is not dependent on a species-specific organ. Rather, the predator's life form is *multiply realizable* across species boundaries. The (morpho)logical independence of the predator-prey relation with respect to species form is especially evident when the relations belong to different genera, such as a praying mantis (predator; insect) catching a small mouse (prey; mammal).

The personal life form is morphologically distinguished from other life forms such as predator or prey by the fact that it is not oriented towards another life form that is different from it. In contrast to other life forms of natural species, the personal life form is distinguished by the fact that it stands in a reflexive difference to the species form to a much higher degree. It is not directed toward other forms of life, but *toward itself*. This self-reflexivity of the personal life form can be further interpreted in a normative sense, as will be shown below, in contrast to species-specific or purely anthropocentric approaches, which are burdened by the is-ought problem or the naturalistic fallacy. The basic intentionality of personhood directed toward other persons opens up a normatively structured space that can be further interpreted as a person space.

The morphology of the personal life form allows us to bring questions of identity and questions of relations into a systematic unity and, in this context, to touch upon questions of the normativity of the personal life form, which are not subject to a naturalistic fallacy. For the personal life form does not concern natural species as such, but higher-level structures and relations that are realized in them. Hence, the complex relation between species and life opens a middle path between personalism and animalism. The personal life form is not to be equated with a certain natural species, but it presupposes natural species in which it formally manifests itself. This means that different biological species, as the human being is one of them, can instantiate the personal life form, which is morphologically qualified by form aspects, independently of the specifications of the respective natural species. These form aspects can be further interpreted in terms of of higher level personal properties, capabilities, and relations. They are normatively assessable, but always remain back-bound to the individual life. Therefore, it must be shown *how* this relation of human species and life form is exactly constituted, especially if the person is also to have normative relevance. What distinguishes persons is not something that belongs to them as a faculty, a property, or a bundle of properties. What distinguishes persons from other living beings is nothing *beyond* life, but something *about life itself*. Therefore, the person must neither be thought as detached from her life, nor be identified with her biological life.

5 Composition

Taking into account the difference between species form and life form, we can apply this distinction to van Inwagen's and Eric Olson's theory of composition.⁴⁵ Van Inwagen describes the relationship between material particles and lives as follows:

($\exists y$ the x s compose y) if and only if the activity of the x s constitutes a life (or there is only one of the x s)⁴⁶

Drawing on van Inwagen's theory of composition, my argument goes as follows: Not only are lives – in the sense of the species form – composed by the activity of material particles, but also – when it comes to particularly complex compositions – persons. The personal composition takes place in two stages. The personal life form is *infigured* in the individual life under the species form. Olson's question, “[u]nder what circumstances do particles compose something other than a mass?”⁴⁷, can thus also be extended to persons, and the conditions necessary for this can be specified.

The systematic advantage of the composition relation is obvious: In contrast to the constitution relation, the composition relation between material particles and a person consists in strict numerical identity, so that the “too many thinkers problem” does not arise here. Therefore, the composition relation can be understood in the sense of an *identity relation*, although it operates with ontologically quite different things.⁴⁸ Like the constitutional relation, the compositional relation is transitive and asymmetrical.⁴⁹ Baker's constitutionalism conceives of the relation of organism and person as that of a “unity without identity”, which raises the question of their coincidence. Baker understands the relationship between person and living organism as being analogous to the relationship between statues and their material. Both – person and statue – are constituted by matter without being identical with it.⁵⁰ However, this analogy is flawed: Persons are mereologically not mere aggregates of particles, but highly complex organizations formed by a higher-level form of life.

In contrast to Baker's constitutionalism, which considers the ontological and moral status of the person independently of the concept of life but only by reference to self-consciousness, I will develop a concept of life that allows the person to be understood as a particular *form* of life. The position of animalism, by its sole concentration on the individual organism, will prove to be an *underdetermined* form of life, which is not able to conceptually distinguish between different forms and stages of life – personal and non-personal. A complex, differentiated, and formal concept of life proves to be a suitable basis to solve the respective problems of animalism and constitutionalism. Life is to be understood not so much materially, as animalism and constitutional theory do, but formally. This formal structure of life allows it to be interpreted in personal terms through relational enrichment and complexity, and to include its ontological as well as moral dimensions.⁵¹

On closer inspection, van Inwagen's "composition question", which allows composition only in the case of life, proves to be inherently undifferentiated, as he understands persons as material objects and argues that they have no immaterial properties.⁵² Van Inwagen's rejection of immaterialism, such as the one advocated by Swinburne,⁵³ however, by no means exhausts all ontological options for determining the person. For it is possible to further distinguish the relation of composition with respect to life, in a way that takes into account the species form and life form of the person. For this purpose, let us consider again the "master argument" of animalism:

- (P1) Presently sitting in your chair is a human animal.
- (P2) The human animal sitting in your chair is thinking.
- (P3) You are the thinking being sitting in your chair.
- (C) Therefore, the human animal sitting in your chair is you.⁵⁴

According to the composition relation, I as a person am identical with the living being sitting on the chair and am not constituted by it in such a way that I am different from it in my personal identity conditions. But here the concept of the "human animal" is morphologically underdetermined. Further formal distinctions – its personal form of life – can be infigured into its species form without having to change its identity conditions. This means, however, that the relation of composition is underdetermined as long as it is not evident which form – the species form or the life form – is materially realized. The relation of composition must therefore be extended by a second-level morphology, so that it is clearly determined under which form the respective composition of matter takes place. Especially in the case of persons who are composed by species form and life form, this question is of crucial importance. Therefore, we need to determine the complex morphology of the person under the relation of configuration or infiguration.

6 Configuration

If the person is to be determined as a living being, but without having to follow the position of animalism and to assume mere biological identity, nor to separate it logically from its bodily organism by a relation of constitution, we need to integrate further determinations into the individual life composed under the species form while preserving the principle of identity. In what follows, I will call this conceptual operation *configuration* or *infiguration*. Configuration means the personal formation of the individual life under the species form, infiguration means the integrative aspect of this configuration. However, it is crucial that the second-order form of life, which is inscribed in the individual human life, is nothing individual. Rather, the personal life form must be determined in such a way that it has general and relational features, which can then, in a second step, be further normatively determined

in such a way that a personal space can be developed, within which persons exist in a normative way.

This means that the composition of the human person must be configurational in two stages, the unity of these stages being developed in such a way as to avoid ontological dualisms:

- 1 Composition of the individual human life under the species form.
- 2 Configuration or infiguration of the personal life form into the human life composed under the species form.⁵⁵

This can be formalized as follows:

- 1 Material composition of human life under the species form as first-order form:
 $(\exists y, \text{the } x\text{s compose } y \text{ biologically})$ if the activity of the $x\text{s}$ composes a human life (where the $x\text{s}$ are material particles).⁵⁶
 and
- 2 Configuration or infiguration of the personal life form at the individually composed life under the species form as a second-order form:
 $(\exists y, \text{the } f \text{ configures/infigures } y \text{ biomorphically})$ if f aligns the capacities of y (i) reflexively and (ii) intersubjectively (where f is the personal life form and y is the human life individually composed under the species form).

Here, we must first explain in more detail what exactly is meant by (i) reflexive and (ii) intersubjective in the context of the personal life form.

Concerning (i): The personal life form relates reflexively to the individual human life under the species form insofar as it refers to it and its capacities and, as it were, starts from it. It being a second-order form ensures that the personal life form is formally independent of the specific qualities of the species form. The personal life form is not exclusively infiguratively dependent on a certain species, but is applicable across species. This reflexive relation to the life composed under the species form can be described in such a way that the person is not only identical with the individual life under the species form, but also *leads* this life.⁵⁷

Re (ii): This reflexive, formal reference to the life composed under the species form is realized in such a way that in it the capacities are intersubjectively oriented in formal independence of the specifications of the species form. In the combination of reflexivity and intersubjectivity, while preserving identity with individual life, a species-spanning form is thus established, which structurally corresponds to what Kant has called “humanity in his person”⁵⁸.

It is decisive that in the configuration of the person the two forms – human species and personal life – are integrated into each other in such a way that

the same material particles do not compose two different objects – a human being *and* a person – which would lead to the “too many thinkers problem” as in the case of the constitution relation. The personal life form is rather an internal unity formation *at* the individual life under the species form of the human being, so that the life form in it further determines the species form. While the species form composes an individual, the life form infigures a supra-individual structure in it, which can be axiologically interpreted as a person space. The personal life form, then, is not to be understood as representing a separate form alongside the human species form, but rather as, by reflexively referring to it, gaining determinations from it. The life form thus represents a kind of internal framework that connects existing capacities under the species form in itself in a specific, i.e., reflexive and intersubjective, way. The composition of the person ensures that the identity with the living organism under the species form is preserved and that the “too many thinkers problem” does not arise as implied by a “unity without identity”. Configuration ensures that the relation of human species form and personal life form is not determined as a mere substitution relation that leaves the relation of the two unresolved, but rather brings the life under species form into an inherently differentiated relation of infiguration with the personal life form. Thus, the configuration/infiguration relation represents an alternative beyond additive and transformation models: It does not add the personal life form onto the species form so that a layered model of capacities is established, nor does it transformatively replace them. Rather, it brings the capacities of the individual life under the species form into such relations that they become actualizable in a new – normative and intersubjective – way.

7 The normativity of life form

However, the question still remains how exactly normativity can be gained from the personal life form. So far, it has merely been shown so far only that it cannot be developed from the species form without committing a naturalistic fallacy. Two structures of the personal life form prove to be particularly central to a normative infiguration of the species form: (i) Reflexivity and (ii) intersubjectivity. These infigurative structures can be considered normatively relevant insofar as they (i) establish a morphological difference from the species form, and (ii) open up a dimension of commitment that is independent of the self-referential, individual interests of the species form. The personal life form is thus neither congruent with life under the species form, nor is it wholly independent of it, but rather it infigures that life in such a way as to confront it with alternative modes of realization of its capacities. This confrontation, which is not dualistically but internally realized, is in turn qualified to contain the character of intersubjectivity insofar as it expresses a universality and generality that transcends the immanent specifications of the life of the species form. This formal dimension of intersubjectivity and generality will be further explained below by reference to Kant’s theory of

rationality and, in particular, his various formulations of the moral law. It will be shown that the normativity of the personal life form springs from a specific interference with the human species form.

Kant fundamentally distinguishes two different perspectives on man, which can be further interpreted in terms of the distinction between human species form and personal life form: Man “as a *sensible being* (*Sinnenwesen*) [...] (belonging to one of the animal species)” and man “as a *being of reason* (*Vernunftwesen*) (not merely a rational being (*vernünftiges Wesen*), because reason, according to its theoretical capacity, could probably also be the quality of a living corporeal being)”.⁵⁹ Kant thus argues that the faculty of reason belongs to man as a natural being, i.e., to the species form, but also, though in a different respect, to man as a person, i.e., to his life form. Kant succeeds in maintaining this distinction and also in motivating it normatively by ensuring that both the law of nature, which shapes the empirical character (life under the human species form), and the law of reason, which determines the personality (life under the personal life form), coexist without contradiction. Kant makes this attempt in the resolution of the Third Antinomy of the *Critique of Pure Reason*. Only through the relatedness of species form and life form, not already through their difference, Kant's theory of normativity receives its full meaning, which purely constitutionalist and coherentist interpretations fail to grasp.

The distinction between human species and personal life forms is substantiated in Kant's theory of the practical use of reason. Kant distinguishes practical intentionality into skill (*Geschicklichkeit*) and morality (*Sittlichkeit*), “the former for arbitrary and accidental, the latter, however, for absolutely necessary purposes”, with the former being “hypothetically necessary”.⁶⁰ Man as a “rational *natural being*” or “the sensible man endowed with reason”⁶¹ is determined by hypothetical imperatives, i.e., individual ends, “and here the concept of an obligation (*Verbindlichkeit*) does not yet come into consideration”.⁶² Kant distinguishes man “according to his *personality*”⁶³ from this species form, and determines him as “a being capable of obligation and indeed considered against himself (humanity in his person)”,⁶⁴ which is under the form of the moral law. The faculty of reason is thus once realized or more precisely *used* only individually-specifically, another time universally-personally. It is, morphologically spoken, both times the *same* faculty, only differently *configured*, namely in the case of the personal life form by reflection and intersubjectivity.

Thus, the personal configuration of reason does not constitute the person, which would then be identity-logically different from the empirical human being. Rather, the personal life form manifests itself in the use of the capacities of life under the species form. This different use of the capacities shows that the personal life form does not separate life under the species form, but that it is always reflexively related to it, indeed interferes with it. In the case of the personal life form, reflexivity as a relation of form is distinguished by the fact that it not only distinguishes itself from the species form and relates to other forms of life, but rather to *itself*, which Kant calls “obligation [...]”

against himself (humanity in his person)”⁶⁵. However, Kant’s talk of “humanity” must not be understood at all in terms of the genealogical species form. Rather, “humanity” denotes the personal life form, since Kant elsewhere also determines humanity as the “personality independent of physical determinations”.⁶⁶ This distinction enables Kant to normatively determine humanity without being guilty of speciesism or a naturalistic fallacy.

We can now look at Kant’s various forms of the categorical imperative as an example for the normative explication of the personal life form, and examine how this interferes with the species form or the empirical character of the human being and becomes normatively significant in it. The focus will be primarily on Kant’s concept of reason and will, in order to show how the personal life form configures these two capacities in normative terms.

The interference of human species and personal life form can be explained using the example of the universalization of maxims, which is morphologically different from hypothetical imperatives according to the distinction of human species and personal life form, and enables normativity by (i) reflexivity and (ii) intersubjectivity.⁶⁷ A well-known formula of the categorical imperative reads: “Act in such a way that you use humanity, both in your person and in the person of everyone else, at all times simultaneously as an end, never merely as a means”⁶⁸. Kant determines this form as “the supreme restrictive condition of all subjective ends” and emphasizes that it ought to be distinguished from the structure of the human species form because of (i) its generality and (ii) its non-empirical but objectivist understanding of humanity.⁶⁹ In the “restrictive condition of all subjective purposes”, it is precisely the confrontational interference of the personal life form with life under the species form that is determined by hypothetical imperatives and thus intersubjectively-universally extended.

In analogy to the twofold formatting of the human rational faculty, Kant distinguishes the faculty of will into a lower and an upper one, which in these two modes can also be related to the human species and personal life form. To distinguish between the two modes, Kant sets up a martial thought experiment that illustrates well the interference of human species form and personal life form.⁷⁰ At the level of immediate action-oriented first-order desires, he identifies the conceivably greatest inclination, namely the survival instinct, or, as Kant puts it, the “love of life”⁷¹ – as the maximum sum of our mere immediate object-oriented preferences, which in their totality constitute the interest of our natural existence and thus concern life under the human species form. Compared to this survival instinct, any particular preference – for instance, in the sense of a “voluptuous inclination” which man claims to be “quite irresistible” – must appear as irrelevant.⁷² Now Kant construes a case in which a person is faced with the choice of either having to “corrupt” an innocent other person “under apparent pretexts”⁷³ or being hanged on the gallows on the spot. Here, according to Kant, it is shown that the person placed before the choice actually still has an alternative to the preferences of the lower faculty of desire, that this natural survival instinct – under the

determination of the species form – is thus precisely not the absolute criterion against which all other preferences must turn out to be lower-valued. Through the form of the moral law – understood as a personal form of life – we are able to take up an epistemic access to an entirely different order than nature, that is, to the normative realm of morality or to “a supersensible nature”⁷⁴, which, according to Kant, constitutes the normative realm of our second-order volitions.⁷⁵ This level of second-order volitions opens up the scope of moral deliberation to the empirically sensible afflicted volitional agent: “He thus judges that he can do something, therefore because he is conscious that he ought to, and recognizes in himself the freedom that would otherwise have remained unknown to him without the moral law”⁷⁶. Human species form and personal life form thus interfere insofar as the latter allows the human faculties of will and reason to align in a way that is reflexive and at the same time intersubjective-universal. As such, the personal life form contrasts with life under the human species form. In that man's capacities are determined by the personal life form, he acts not as an individual human life determined by hypothetical imperatives, but as a person who can generalize his interests supra-individually through the use of reason and the upper faculty of desire. Kant describes this as follows: “Accordingly, every rational being must act as if, through its maxims, it were at all times a legislative member in the general realm of ends”⁷⁷.

In its normativity, the personal life form is not to be understood in such a way that only *moral* actions can be realized under it. Rather, immoral actions are also compatible with the personal life form in that human capacities can be realized *under* but at the same time *against it*. The personal life form does not lose its normativity even when immoral actions are performed under it. Rather, the personal life form is the condition of possibility for actions to become morally qualifiable at all.

8 Conclusion

I have argued that a unified concept of the person can be consistently developed, one which contains both theoretical-ontological and normative-practical dimensions and which manages to make them further comprehensible. I have used the concept of the personal life form as the guiding concept of that explication, and the difference between (human) species form and (personal) life form served as the basic morphological distinction. In particular, my goal was to define the concept of life unambiguously, beyond merely metaphorical uses, as in the distinction between “first” and “second nature”. Therefore, an undisputed biological concept of life was taken as a starting point, but which very soon conceptually pointed beyond itself, in that the concept of the life form came into view as a systematic starting point of a genuinely philosophical morphology. Insofar the personal life form is infiguratively related to the individual human-biological life, and while preserving the strict numerical principle of identity, it proves to

be equally biologically and axiologically comprehensible. It allows to consistently think a transition from natural states and entities, such as life, to normative realities, such as persons, without succumbing to speciesism or naturalistic fallacy.

Notes

- 1 See Baker (2008, 10), who raises the question of “the kind of beings that we fundamentally are”.
- 2 Schechtman (2014), 2.
- 3 Olson (2016).
- 4 In doing so, I take up Buchheim/Noller (2016), but I explicate the concept of personhood in a different sense by further differentiating the concept of life.
- 5 For example in Foot (2004, 161), where “life form” is referred to “species” in the index. This has been pointed out by Toepfer (2020, 3). Toepfer explicitly distinguishes between species and life form, though he does not further determine their relation in terms of a theory of personal identity.
- 6 See Olson (2003), 320: “‘Animalism’ is [...] the view that we are essentially or most fundamentally animals. We are essentially animals if we could not possibly exist without being animals”.
- 7 See Baker (2008), 10.
- 8 See Baker (2000), 98: “[W]hen an animal has a first-person perspective or the capacity for one, a new entity comes into being: a person”.
- 9 Baker (2016), 56.
- 10 Baker (2002b), 387.
- 11 For contemporary theories of substance dualism, see Swinburne (1997), 145:

I understand by substance dualism the view that those persons which are human beings (or men) living on Earth, have two parts linked together, body and soul. A man’s body is that to which his physical properties belong. If a man weighs ten stone then his body weighs ten stone. A man’s soul is that to which the (pure) mental properties of a man belong. If a man imagines a cat, then, the dualist will say, his soul imagines a cat.
- 12 Baker (2016), 53.
- 13 Shoemaker/Strawson (1999), 291.
- 14 Parfit (2012), 7.
- 15 See Zimmermann (2008, 26), who speaks of a “Master Argument for Animalism”.
- 16 Blatti (2020).
- 17 Baker (2002), 33.
- 18 Baker (2016), 58.
- 19 Baker (2008), 10.
- 20 Shoemaker (2016), 303 f.
- 21 See Schechtman (2014), 198: “[W]e can see persons as complicated biological-psychological-social loci more fundamental than the animalist’s ‘organisms,’ which are viewed as abstractions”.
- 22 Olson (1999), 165.
- 23 A life (form)-centered approach to determining the person is undertaken, in various forms, by Spaemann (1996), Thompson (2011), Quante (2015), Schechtman (2014), Buchheim (2019), and Wunsch (2019).
- 24 See Toepfer (2020), 3.
- 25 Toepfer (2011), 484.
- 26 Toepfer (2011), 484.

- 27 See on the relationship between species form and life form Buchheim/Noller (2016), 164 f.
- 28 A similar concept is that of the “biographical”, which can be found in Buchheim (2019). However, the term “biomorphic” seems to me to be more suitable for the concept of the personal life form because it avoids confusion with purely cultural techniques from the outset and establishes even more strongly the reference to the species, in its simultaneous formal difference.
- 29 Here the question arises further, how the individual life relates to its kind form. In the following, when I speak of the (human) species form, I will always mean also the individual (human) life, which falls under it or forms it.
- 30 See Buchheim/Noller (2016), 149.
- 31 On the notion of property as a “distinguishing and exclusive characteristic of living things,” which holistically concerns the *whole* living individual, i.e., is *grounded* in life, see Buchheim (2014), 174 and Buchheim (2016), 22.
- 32 Kant, MS, 6:418.
- 33 Kant, MS, 6:418.
- 34 It should be noted here that Kant understands “humanity” in the *person* of man not in the sense of the biological species, but in the sense of the personal form of life – a distinction that is often not made and is thus the source of naturalistic fallacies.
- 35 Kant, GMS, 4:408. The structural proximity of Kant’s a priori moral law to the (logically understood) form of life has been pointed out by Thompson (2008, 6).
- 36 On the independence of species form and life form, see Buchheim/Noller (2016), 149. In the following, on the other hand, I will argue more strongly for the connection between species form and life form in an inherently differentiated concept of life, and apply this to the concept of person in the context of a complex morphology.
- 37 Kant determines this general structure as
 - systematic connection of rational beings by common objective laws, i.e. [as] a realm, which, because these laws have the very relation of these beings to each other as ends and means as their intention, can be called a realm of ends [...].

(GMS, 4:433)
- 38 See Buchheim/Noller (2016), 164.
- 39 Buchheim (2019) rightly points to the intrinsic-relational constitution of persons while preserving their numerical identity, following Spaemann (1996).
- 40 Schechtman (2014).
- 41 Spaemann (1996), 78.
- 42 McDowell (1996).
- 43 For the concept of speciesism see Singer (2002), 6: “Speciesism [...] is a prejudice or attitude of bias in favor of the interests of members of one’s own species and against those of members of other species”.
- 44 See, for example, Foot (2001).
- 45 van Inwagen (1990); Olson (2007), 231.
- 46 van Inwagen (1990), 82.
- 47 Baker (1999).
- 48 See Lewis (1991), 82:

I say that composition – the relation of part to whole, or, better, the many-one relation of many parts to their fusion – is like identity. The ‘are’ of composition is, so to speak, the plural form of the ‘is’ of identity. Call this the Thesis of *Composition as Identity*.

- 49 See Doepke (1997), 19.
- 50 Baker (2007), 266: “According to the Constitution View [...] human persons are wholly constituted by human bodies (= human animals), just as marble statues are wholly constituted by pieces of marble”.
- 51 For a similar approach, albeit disregarding the infigurative dimension of the individual living person, see Buchheim (2019), 49 ff.
- 52 van Inwagen (2007), 206.
- 53 Swinburne (1997).
- 54 Blatti (2020).
- 55 The term “installation” in the sense of “setting in place” would also be suitable for this step, but I prefer it to the term “infiguration” because of its technical connotations.
- 56 van Inwagen (1990, 82) speaks of “constitute” here. However, mereologically this does not mean Lynne Baker’s relation of constitution. Rather, “constitute” here must be understood more unspecifically, in the sense of “to generate” or even “to form”.
- 57 Plessner (1981, 391) speaks of man as a person “leading” a life. See also Spaemann (1996, 260), who distinguishes between “being” and “having” a nature. See further Schechtman (2011), 395 and Buchheim (2019, 39), who speaks of the “biographical existence” of persons in contrast to their “biological existence”.
- 58 Kant, KpV, 5:87.
- 59 Kant, MS, 6:418.
- 60 Kant, KrV, B 851.
- 61 Kant, Rel., 6:439 fn.
- 62 Kant, MS, 6:418.
- 63 Kant, MS, 6:418.
- 64 Kant, MS, 6:418.
- 65 Kant, MS, 6:418.
- 66 Kant, MS, 6:239.
- 67 See Noller (2016), 154–156.
- 68 Kant, GMS, 4:429.
- 69 Kant, GMS, 4:431.
- 70 I have also interpreted this example in the context of Kant’s theory of autonomy. See Noller (2016), 152 f.
- 71 Kant, KpV, 5:30.
- 72 Kant, KpV, 5:30.
- 73 Kant, KpV, 5:30.
- 74 Kant, KpV, 5:43.
- 75 Here I follow the terminology of Frankfurt (1971).
- 76 Kant, KpV, 5:30.
- 77 Kant, GMS, 4:438.

Bibliography

- Baker, Lynne R. (1999): “Unity without Identity: A New Look at Material Constitution”, in: *Midwest Studies in Philosophy* 23/1, 144–165.
- Baker, Lynne R. (2000): *Persons and Bodies. A Constitution View*, Cambridge: Cambridge University Press.
- Baker, Lynne R. (2002a): “On Making Things Up: Constitution and Its Critics”, in: *Philosophical Topics* 30/1, 31–51.
- Baker, Lynne R. (2002b): “The Ontological Status of Persons”, in: *Philosophy and Phenomenological Research* 65/2, 370–388.

- Baker, Lynne R. (2007): "Persons and the Natural Order", in: *Persons. Human and Divine*, ed. Peter van Inwagen u. Dean Zimmermann, New York: Oxford University Press, 261–278.
- Baker, Lynne R. (2008): "Big Tent Metaphysics", in: *Abstracta*, Special Issue I, 8–15.
- Baker, Lynne R. (2016): "Animalism vs. Constitutionalism", in: *Animalism: New Essays on Persons, Animals, and Identity*, ed. Stephan Blatti u. Paul F. Snowdon, New York: Oxford University Press, 50–63.
- Baker, Lynne (2019): "What Are Persons, and How Do They Exist?", in: Jörg Noller ed.: *Was sind und wie existieren Personen? Probleme und Perspektiven der gegenwärtigen Forschung*, Leiden/Boston, MA: Mentis, 15–29.
- Blatti, Stephan (2020), "Animalism", *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), ed. Edward N. Zalta (<https://plato.stanford.edu/archives/fall2020/entries/animalism/>).
- Buchheim, Thomas (2014): "Vermögen als Ursachen", in: Anne Sophie Spann/Daniel Wehinger (eds.): *Vermögen und Handlung. Der dispositionale Realismus und unser Selbstverständnis als Handelnde*, Paderborn: Mentis, 173–189.
- Buchheim, Thomas (2016): "Remarks on the Ontology of Living Beings and the Causality of Their Behavior", in: Miguel Garcia-Valdecasas/José Ignacio Murillo/Nathaniel F. Barrett (eds.): *Biology and Subjectivity. Philosophical Contributions to Non-reductive Neuroscience*, Berlin: De Gruyter, 63–75.
- Buchheim, Thomas (2019): "What Are Persons? Reflections on a Relational Theory of Personhood", in: Jörg Noller (ed.): *Was sind und wie existieren Personen? Probleme und Perspektiven der gegenwärtigen Forschung*, Leiden/Boston, MA: Mentis, 31–55.
- Buchheim, Thomas/Noller, Jörg (2016): "Sind wirklich und, wenn ja, warum sind alle Menschen Personen? Zu Robert Spaemanns philosophischer Bestimmung der Person", in: *Die Person – ihr Selbstsein und ihr Handeln. Zur Philosophie Robert Spaemanns*, ed. Josef Kreiml u. Michael Stickelbroeck, Regensburg: Pustet, 145–179.
- Doepke, Frederick C. (1997): "Spatially Coinciding Objects", in: Michael C. Rea (ed.): *Material Constitution. A Reader*, Lanham, MD: Rowman & Littlefield, 10–24.
- Foot, Philippa (2001): *Natural Goodness*, New York: Oxford University Press.
- Foot, Philippa (2004): *Die Natur des Guten*, übersetzt von Michael Reuter, Frankfurt/M.: Suhrkamp.
- Frankfurt, Harry G. (1971): "Freedom of the Will and the Concept of a Person", in: *The Journal of Philosophy* 68/1, 5–20.
- Kant, Immanuel (1900 ff.): *Akademie-Ausgabe*, ed. Preußische Akademie der Wissenschaften, Berlin: De Gruyter (all translations are mine).

GMS Grundlegung zur Metaphysik der Sitten

KpV Kritik der praktischen Vernunft

KrV Kritik der reinen Vernunft

MS Metaphysik der Sitten

Rel. Die Religion innerhalb der Grenzen der bloßen Vernunft

Lewis, David (1991): *Parts of Classes*, Oxford: Blackwell.

McDowell, John (1996): *Mind and World*, Cambridge/London: Harvard University Press.

Noller, Jörg (2006): *Die Bestimmung des Willens. Zum Problem individueller Freiheit im Ausgang von Kant*. Freiburg/München: Alber.

Olson, Eric T. (1999): "Reply to Lynne Rudder Baker", in: *Philosophy and Phenomenological Research*, 59/1, 161–166.

- Olson, Eric T. (2003): "An Argument for Animalism", in: *Personal Identity*, ed. Raymond Martin u. John Barresi, Oxford: Wiley-Blackwell, 318–334.
- Olson, Eric T. (2007): *What are We. A Study in Personal Ontology*, New York: Oxford University Press.
- Olson, Eric T. (2016), "Personal Identity", in: *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), ed. Edward N. Zalta (<http://plato.stanford.edu/archives/spr2016/entries/identity-personal>)
- Parfit, Derek (2012): "We Are not Human Beings", in: *Philosophy* 87/1, 5–28.
- Plessner, Helmuth (1981): *Die Stufen des Organischen und der Mensch. Einleitung in die philosophische Anthropologie*, in: Ders.: *Gesammelte Schriften IV*, ed. Günter Dux, Odo Marquard u. Elisabeth Ströker, Frankfurt a.M.: Suhrkamp.
- Quante, Michael (2015): "Pragmatistische Anthropologie und Ethik in Anwendung: eine philosophische Skizze", in: *Anthropologie und Ethik*, ed. Jan-Christoph Heilinger u. Julian Nida-Rümelin, Berlin/Boston, MA: De Gruyter, 37–58.
- Schechtman, Marya (2011): "The Narrative Self", in: Shaun Gallagher (ed.), *The Oxford Handbook of the Self*, New York: Oxford University Press, 394–416.
- Schechtman, Marya (2014): *Staying Alive. Personal Identity, Practical Concerns, and the Unity of a Life*, New York: Oxford University Press.
- Shoemaker, David (2016): "The Stony Metaphysical Heart of Animalism", in: Stephan Blatti/Paul Snowdon (eds.), *Animalism*, New York: Oxford University Press, 303–328.
- Shoemaker, Sydnes/Strawson, Galen (1999): "Self and Body", in: *Proceedings of the Aristotelian Society*, Supplementary Volumes 73, 287–332.
- Singer, Peter (2002): *Animal Liberation*, New York: HarperCollins [¹1975].
- Spaemann, Robert (1996): *Personen. Versuche über den Unterschied zwischen 'etwas' und 'jemand'*, Stuttgart: Klett.
- Swinburne, Richard (1997): *The Evolution of the Soul*. Revised Edition, New York: Oxford University Press.
- Thompson, Michael (2008): *Life and Action. Elementary Structures of Practice and Practical Thought*, Cambridge/MA: Harvard University Press.
- Thompson, Michael (2011): *Leben und Handeln. Grundstrukturen der Praxis und des praktischen Denkens*, Übers. von Matthias Haase, Berlin: Suhrkamp.
- Toepfer, Georg (2011): Art. "Lebensform", in: Toepfer, Georg (ed.): *Historisches Wörterbuch der Biologie. Geschichte und Theorie der biologischen Grundbegriffe*. Bd. 2, Stuttgart: Metzler, 484–496.
- Toepfer, Georg (2020): "Der Begriff der Lebensform als typologisches, nicht-normatives Konzept der Biologie", in: Martin Hähnel/Jörg Noller (eds.): *Die Natur der Lebensform. Perspektiven in Biologie, Ontologie und praktischer Philosophie*, Leiden: Brill, 3–16.
- van Inwagen, Peter (1990): *Material Beings*, Ithaca/London: Cornell University Press.
- van Inwagen, Peter (1997): "Materialism and the Psychological-Continuity Account of Personal Identity", in: *Philosophical Perspectives* 11, 305–319.
- van Inwagen, Peter (2007): "A Materialist Ontology of the Human Person", in: P. van Inwagen/Dean Zimmermann (eds.): *Persons. Human and Divine*, Oxford: Oxford University Press, 199–215.
- Wunsch, Matthias (2019): "Personalität, Würde und Lebensform", in: Jörg Noller (ed.): *Was sind und wie existieren Personen? Probleme und Perspektiven der gegenwärtigen Forschung*, Leiden/Boston, MA: Mentis, 185–207.
- Zimmermann, Dean (2008): "Problems for Animalism", in: *Abstracta* 4, Special Issue I, 23–31.

Index

Italic page numbers refer to figures and page numbers followed by “n” denote endnotes.

- abstract entities 59–60, 62
- abstract extensibility 74
- abstraction 52, 59–62
- activity 45–47, 49
- affective associations 138
- affective binding 140–141
- agent 84, 87, 89, 91–93, 127, 135, 139, 165–166, 201–202
- agential approaches 1
- animal 5–9, 17, 40, 42, 44, 47–51, 104, 106, 151, 160, 209, 213
- animalism 29–30, 40–42, 69, 100–123, 150, 159, 208–211, 215; PLV and 159–162
- animalist ontology 101, 117; *versus* psychological ethics 101–105
- animality 5–6
- animals 5–11, 17, 40–42, 47, 49, 51, 100, 106–107, 160, 208, 209
- Aristotle 68–69
- autobiographical memory 136–137
- autopoiesis theory 116
- autopoietic system 112

- Baker, L. R. 5, 10–12, 17, 21, 21n2, 36–37, 121n5, 208–210, 215, 224n50; metaphysical significance, personhood 10–12; principle, metaphysics of value 12–13; principle, problems 13–15
- Baker's principle of the metaphysics* 13
- Baylis, F. 182n19
- Beckner, Morton 142n1
- bio-agency 113–114
- biographical pattern, of representation 75–76
- biographical situation 75

- biological/biology 36, 68, 70–72, 100–101, 103–105, 108–109, 112–114, 117–120, 150–151, 158–162, 207, 209–211; approach 100, 103–104, 150–151; continuity 23–24, 28, 101, 103–105, 117, 150–152; identity 100, 113–114, 118, 122n19, 216; life 20, 70–72, 101, 153, 214; species 208, 212, 214; subjectivity 100–123; substances 101, 108–109
- Biological Continuity Criterion 101–102
- body 23, 26–28, 42, 46, 102, 104, 134, 138, 162, 169, 173–174, 209
- Bowne, Borden 64n5
- brain 8–9, 15–17, 20, 41–42, 102, 109, 130, 151–152, 159–160; transplant 8, 41–42, 152; transplanted 17
- Brainy 16–17
- Bratman, Michael 82–84, 87, 90–91, 97n35
- Brown, Borden 53
- Bruner, J.S. 135–136
- Buchheim, Thomas 223n28

- capacity preservation 50
- Capra, Frank 200
- Cavell, Stanley 199–200
- Chiong, Winston 160–161
- Cocking, D. 186
- cognition 101, 109–112, 114, 116, 158
- cognitive changes 141
- coherence: in self-pattern 127–143; unity and 130–134
- collective intentionality 185–203
- common heritage of life 71

- commonsense view 56
- communication 198–202
- composition 215–216
- configuration 216–218
- conscious/consciousness 1, 6–7, 9, 13–15, 25–26, 28, 31–32, 35, 58, 64, 69–70, 81, 93, 109–112, 114, 128, 138, 188, 208–209
- constitution theory 55
- constitution view/constitutionalism 133, 193–194, 208–211, 215
- corporate we-identities 187–191; robustness of 190
- criminal identity 167–168, 176–177
- cumulative network model (CNM) 168–170, 175–177, 179
- cumulative network self 169–174
- cumulative process 172

- de dicto* necessities 27
- Dennett, D.C. 128, 138
- dependence model 1
- depression 141
- Dewey, John 128
- Diamantis, M. 167, 175–176
- dualism: Cartesian 15; of mind 20

- essentialism 29; strong Lockean person 31; strong person 28, 30; weak person 28, 30–31; weak psychological 35

- Ferrero, Luca 96n13
- fiction 138–141
- filiation 72; of life 69–70
- 4d-structures 54–55
- four-dimensionalism 18
- Frege, Gottlob 60–61, 63
- friendship 185–189, 191–193, 195, 197–202; constitution 185–203; depth of 198–202; historicity of 199
- Frith, C.D. 139–140

- Gallagher, S. 195–197
- genidentity 54–55, 58, 60, 62–64, 65n6
- Ginsburg, Simona 110
- Godfrey-Smith, Peter 114, 123n32
- group membership 186
- Gurwitsch, Aron 198

- Hale, Bob 60–61
- Hawthorne, Nathaniel 149
- historicity 198–202
- Homo sapiens* 5
- human 5–6, 8–9, 44, 47–50, 106–107, 151, 155, 158–159, 161–162, 217, 220–221
- human cultures 156–158
- humanity 212, 217, 219–220
- Hume, David 53, 131–132, 138
- Hurley, Susan 123n30
- hypothetical cerebrum transplantation 120

- idem identity 132
- identification 174–175
- identity 15–21; meanings of 168; numerical 17, 19, 54, 58, 70, 100, 103–104, 107, 117, 119
- immanent causation 60
- individual life 212–218
- individuals: multiplicity 69–70
- intelligence 15
- intentions 87; commitment 82–84
- interpersonal relations 198–199, 201–202
- intersubjective interactions 137
- intersubjectivity 218, 220
- ipse identity 132, 134, 142
- isotopic life-association 68, 73, 75
- isotopy 72–74
- It Happened One Night* 200
- I-we relation 185, 189

- Jablonka, Eva 110
- James, W. 129
- Jonas, Hans 116–117, 123n29

- Kant, Immanuel 212, 218–221, 223n37
- Kelso, Scott 127
- Kennett, J. 186
- Kiverstein, J. 136
- Korsgaard, Christine 84, 97n22, 98n36
- Kyselo, Miriam 130

- Layton, A.S.F. 130
- Lefty and Righty, brain 17–18, 20
- Lewis, David 56
- life 68–75, 153, 211–221; biographical 70–72, 74–76; biological concept of 221; burden of 71; human 76, 168, 213, 217; of persons 41
- life form 208, 211–221; biomorphic 212
- living beings 69–70, 74, 76, 207, 211, 214
- Locke, John 6, 23, 30, 33, 41, 53, 69–70, 81–82, 87, 138, 150, 166
- Lotze, Hermann 53
- Lowe, E. J. 115
- Lyon, Pamela 110

- MacKenzie, Katrina 142
Madden, Rory 21n3, 42, 51; proposal 42–43; theory 43–44
Martin, C. B. 45
Matsumura, Kaiponanea 167, 178
Maturana, Humberto R. 110, 112
Mayr, Erasmus 96n7
McIntyre, Alistair 136
Meincke, Anne Sophie 121n6
memory 135–137, 168
mental powers 5–8, 11, 14, 18, 35–36
mental states 111
metabolism 113, 116
metaethics 13
metaphysical/metaphysics 6–16, 19–21, 52–55, 59, 63, 150, 160, 210; significance 7, 10, 12, 15
‘metaphysics–first’ proposal 19–20
Metzinger, T. 139–140
mind–life dualism 15
morality 219, 221
moral law 212, 219, 221
moral/morality 6–7, 12–15, 19, 81, 100, 104, 154, 165, 212, 215, 219, 221
moral responsibility 81, 103–104, 133, 150–151, 154, 192
moral significance 6, 12–15, 100
motivation 111
- narratives 92, 129, 132–142, 185–203; approach 187, 201–202; competency 135; framework 135, 137; identity 138–141, 191–195, 199; intimacy 191
narrative self 134
narrative self-constitution view 133, 193–194
Nehamas, A. 201
Neisser, Ulrich 127
Nelson, Katherine 136
Neo-Platonism 59
The Network Self: Relation, Process and Personal Identity 168
nonpeople 7
Noonan, Harold 23, 34
normative/normativity 1–2, 82, 85–86, 88–90, 93–95, 137, 141–142, 210, 213–214, 218–221; commitment 81–82, 85–86, 89, 95, 168; commitment, transfer 85–93; factors 137; of life form 218–221; transfer 82, 85, 88–90, 93–96
No–Yes camp 23
obligation 180
Olson, Eric 29–38
Olson, Eric T. 1, 23–24, 28–29, 34, 36–38, 100–109, 117–119, 121n1, 121n4, 121n11, 123n35, 151–152, 158, 208, 211, 215
ontological/ontology 1–2, 10–11, 36, 52, 54–60, 76, 100–101, 103–105, 107, 115, 117–119, 207, 209–210, 215; approaches 1; framework 108, 115; significance 10–11
organic/organism 8–9, 14–15, 36, 51, 69, 101, 106, 108, 111, 113–114, 116–118, 159–161, 169, 198, 208–210, 215; human 8, 101, 108, 119–120, 161
- Parfit, Derek 23, 84, 97n19, 97n21, 119, 166
Parfit-Shoemaker Thesis 102
permanence 59–61
Perry, John 23
persistence 8–9, 11–14, 16, 19, 21, 24–26, 29, 37, 43–44, 47–51, 106
persistent vegetative state (PVS) 101–102, 119–120, 161
personal identity 1–2, 9, 12, 16, 34–35, 37, 42, 52, 57, 62, 81–82, 85, 87, 92–96, 100, 107, 150, 159, 166–169, 191; cumulative network self and 169–174; examples in law 166–168; internal coherence 53–54; neo-Lockean account of 85–93; problem of becoming 52–53; process account 52–65; responsibility and 165–182; transfer and overlapping processes 93–95; transfer of commitment 81–98
personalism 207–208, 214
personal life 41, 58, 70, 72, 74, 217; biographically realized isotopy 72–74; form 214, 217–218, 221
personal persistence: hybrid account of 23–38
personal we-identities 187–191, 193, 195, 197–198, 202
personhood 5–8, 10–14, 34, 113, 120, 152–153, 155–157; and animality 5–6; metaphysical insignificance of 6–12; primary kind 12, 14
Person Life View (PLV) 149–150, 153–159, 169; animism and 159–162
persons 40–41, 67–77, 92; form 212–214; life 207–224
Persons 74

- person space 149–151, 153, 155,
 157–159, 161–162, 214, 218; animals
 in 149–163; anthropological account
 156–159; biological and psychological
 accounts 150–153; Person Life View
 153–156
 person's traits: clusters 171; network 170
 Plessner, Helmuth 224n57
 plural agent 197, 201, 202
 precommitment device 178
 presentism 56–58
Problem of Personal Persistence 25
 process 52–55, 57–60, 62, 94, 95,
 112–115, 119, 169, 171, 173, 175
Process and Reality 52
 process metaphysics 54–55, 57, 59
 process ontology 52, 54–58, 62;
 endurance of persons 57–59
 processual animalism 100–123
Processual Animalism 101
 Psychological Continuity Criterion 102
 psychological ethics 101; *versus* animalist
 ontology 101–105
 psychological/psychology 23, 27–33,
 35–38, 100–103, 105–108, 118–119,
 150–152, 166, 169; approach 100,
 102–103, 105, 107, 119, 121n4,
 121n5, 150–153; connectedness
 31–32; continuity 8–9, 23–24, 27–28,
 30–31, 35, 37, 42, 102–103, 105, 117,
 150, 166; criteria 166, 169, 175, 178;
 essentialism 35; theorists 26, 29–32,
 34–37, 150–152

 Quinton, A. 180n3

 rational stability 83–84, 96n6
 reactivity 45
 reality 138–141
 reason 6–7, 16–19, 23, 48, 84, 93–94, 158,
 200, 219, 221
Reasons and Persons 84
 reflexivity 220
 Reid, Thomas 166
 re-identification 191–195
 relevant traits 173–174, 176–177,
 179–180
 representation, pattern 74–77
 Rescher, Nicholas 60
 responsibility 81–82, 103–104, 150–151,
 165–182; identity and 165–168
 Ricoeur, Paul 132, 134–136, 139, 142,
 143n3

 Rietveld, E. 136
 “round-game” argument 94
 Russell, Bertrand 57

 Schaffner, Kenneth 142n1
 Schechtman, Marya 1, 97n17,
 132–134, 138–139, 169, 191–193,
 207, 210
 Scheler, Max 191
 Schroer, Jeanine Weekes 92–93
 Schroer, Robert 92–93
 self-conceptions 129, 195
 self-conscious/self-consciousness 5–7,
 10–11, 14–15, 25, 28, 31–32, 35, 37,
 41, 58, 70, 128, 207–211; power of
 10–11
 self-constitution 133
 selfhood 115, 117, 134, 193
 self-identities 130, 137, 187, 191,
 193–195, 201–202
 self-knowledge 127
 self-narratives 132, 134, 136–137, 141,
 193–195, 201–202
 self-patterns 127–135, 137–142;
 affective processes 128; behavioral/
 action aspects 128; bodily existence
 128; cognitive and psychological
 factors 129; ecological relations 129;
 intersubjective factors 128; narrative
 aspects 129; normative factors 129;
 prereflective experiential aspects 128;
 reflective capacities 129
 self-process 170–175, 177, 180
 Sherrington, C.S. 130
 Shoemaker, David 210
 Shoemaker, Sydney 23, 29, 32–34, 36, 42
 Simons, Peter 60
 Snowdon, Paul F. 23, 121n2
 social groups 185–186, 190
 social interactions 128, 149, 158–159
 Spaemann, Robert 67, 74
 species 71–75, 208, 211–221; natural 72,
 211, 214
 speciesism 72, 213, 220, 222
 Strawson, Galen 65n12, 133
 subjectivity 2, 101, 105–109, 111–112,
 114–119; biological 100–123;
 exclusion of 105–109; personhood,
 empirical discoveries 109–112;
 reappearance, processual animalism
 112–117
 substance 37, 52–55, 58–59, 63, 101, 106–
 108, 115, 118, 209; concept 106, 108;

dualism 14; identity 115; ontological
 animalism 105–109; ontology 59
 substratum theory 122n12
 Swinburne, Richard 21n2, 115

 Taylor, Charles 137
 Thompson, Judith J. 23
 Thompson, Michael 67, 77n4
 3d-objects 54–55
 3d-substances 54–55, 58, 63
 Tollefsen, D. 195–197
 transplanted brain 17

 valence 111–112
 value 5–7, 11–16, 19–21, 111, 129, 195;
 theory 20–21
 ‘value-first’ proposal 19–20
 Van Inwagen, Peter 215–216, 224n56
 Varela, Francisco 110, 112, 116
 Varga, Somogy 143n4

Wakefield 149, 162
 Wallace, K. 167, 181n6, 182n23
 Watson, G. 165
 weak person essentialism 28, 30, 31
 weak psychological essentialism 35
 we-identities 185–203; ephemeral,
 corporate and personal 188–191; of
 friendship 201–202; status of
 187, 195
 we-narratives: corporate groups
 195–198
 Whitehead, Alfred N. 52, 54–55, 59–60
 Wiggins, David 43, 121n2
 Williams, Bernard A. O. 23, 42
 Witt, Karsten 24, 28–38
 Wright, Crispin 60–61

 Yes–Yes theory 24

 Zahavi, Dan 123n29